Flamm, W. G. October 13, 1981. Remarks to the Committee on the Institutional Means for Assessment of Risks to Public Health.

Higginson, J. 1969. Present trends in cancer epidemiology. In Proc. Can. Cancer Conf. 8:40-75.

Maugh, T. 1978. Who chooses chemicals for testing? Science 201:1200.

Tomatis, L., C. Agthe, H. Bartsch, J. Huff, R. Montesano, R. Saracci, E. Walker, and J. Wilbourn. 1978. Evaluation of the carcinogenicity of chemicals: a review of the monograph program of I.A.R.C. Cancer Res. 38:877.

I
The Nature of Risk Assessment

Recent criticisms of the conduct and use of risk assessment by regulatory agencies have led to a wide range of proposed remedies, including changes in regulatory statutes and the development of new methods for assessing risk. The mandate to this Committee was more limited. Our objective was to examine whether alterations in institutional arrangements or procedures, particularly the organizational separation of risk assessment from regulatory decision-making and the use of uniform guidelines for inferring risk from available scientific information, can improve federal risk assessment activities.

Before undertaking to determine whether organizational and procedural reforms could improve the performance and use of risk assessment in the federal government, the Committee examined the state of risk assessment and the regulatory environment in which it is performed. In this chapter, we define risk assessment and differentiate it from other elements in the regulatory process, analyze the types of judgments made in risk assessment, and examine its current government context. Because one chronic health hazard, cancer, was highlighted in the Committee's congressional mandate and has dominated public concern about public health risks in recent years, most of our report focuses on it. Furthermore, because activities in four agencies--the Environmental Protection Agency (EPA), the Food and Drug Administration (FDA), the Occupational Safety and Health Administration (OSHA), and the Consumer Product Safety Commission (CPSC)--have given rise to many of the proposals for changes in risk assessment practices, our review focuses on these four agencies. The conclusions of this report, although directed primarily at risk assessment of potential carcinogens as performed by these

four agencies, may be applicable to other federal programs to reduce health risks.

TERMINOLOGY

Despite the fact that risk assessment has become a subject that has been extensively discussed in recent years, no standard definitions have evolved, and the same concepts are encountered under different names. The Committee adopted the following terminology for use in this report.

RISK ASSESSMENT AND RISK MANAGEMENT

We use risk assessment to mean the characterization of the potential adverse health effects of human exposures to environmental hazards. Risk assessments include several elements: description of the potential adverse health effects based on an evaluation of results of epidemiologic, clinical, toxicologic, and environmental research; extrapolation from those results to predict the type and estimate the extent of health effects in humans under given conditions of exposure; judgments as to the number and characteristics of persons exposed at various intensities and durations; and summary judgments on the existence and overall magnitude of the public-health problem. Risk assessment also includes characterization of the uncertainties inherent in the process of inferring risk.

The term risk assessment is often given narrower and broader meanings than we have adopted here. For some observers, the term is synonymous with quantitative risk assessment and emphasizes reliance on numerical results. Our broader definition includes quantification, but also includes qualitative expressions of risk. Quantitative estimates of risk are not always feasible, and they may be eschewed by agencies for policy reasons. Broader uses of the term than ours also embrace analysis of perceived risks, comparisons of risks associated with different regulatory strategies, and occasionally analysis of the economic and social implications of regulatory decisions--functions that we assign to risk management.

The Committee uses the term risk management to describe the process of evaluating alternative regulatory actions and selecting among them. Risk management, which is carried out by regulatory agencies under various legislative

mandates, is an agency decision-making process that entails consideration of political, social, economic, and engineering information with risk-related information to develop, analyze, and compare regulatory options and to select the appropriate regulatory response to a potential chronic health hazard. The selection process necessarily requires the use of value judgments on such issues as the acceptability of risk and the reasonableness of the costs of control.

STEPS IN RISK ASSESSMENT

Risk assessment can be divided into four major steps: hazard identification, dose-response assessment, exposure assessment, and risk characterization. A risk assessment might stop with the first step, hazard identification, if no adverse effect is found or if an agency elects to take regulatory action without further analysis, for reasons of policy or statutory mandate.

Of the four steps, hazard identification is the most easily recognized in the actions of regulatory agencies. It is defined here as the process of determining whether exposure to an agent can cause an increase in the incidence of a health condition (cancer, birth defect, etc.). It involves characterizing the nature and strength of the evidence of causation. Although the question of whether a substance causes cancer or other adverse health effects is theoretically a yes-no question, there are few chemicals on which the human data are definitive. Therefore, the question is often restated in terms of effects in laboratory animals or other test systems, e.g., "Does the agent induce cancer in test animals?" Positive answers to such questions are typically taken as evidence that an agent may pose a cancer risk for any exposed humans. Information from short-term in vitro tests and on structural similarity to known chemical hazards may also be considered.

Dose-response assessment is the process of characterizing the relation between the dose of an agent administered or received and the incidence of an adverse health effect in exposed populations and estimating the incidence of the effect as a function of human exposure to the agent. It takes account of intensity of exposure, age pattern of exposure, and possibly other variables that might affect response, such as sex, lifestyle, and other modifying factors. A dose-response assessment usually

requires extrapolation from high to low dose and extrapolation from animals to humans. A dose-response assessment should describe and justify the methods of extrapolation used to predict incidence and should characterize the statistical and biologic uncertainties in these methods.

Exposure assessment is the process of measuring or estimating the intensity, frequency, and duration of human exposures to an agent currently present in the environment or of estimating hypothetical exposures that might arise from the release of new chemicals into the environment. In its most complete form, it describes the magnitude, duration, schedule, and route of exposure; the size, nature, and classes of the human populations exposed; and the uncertainties in all estimates. Exposure assessment is often used to identify feasible prospective control options and to predict the effects of available control technologies on exposure.

Risk characterization is the process of estimating the incidence of a health effect under the various conditions of human exposure described in exposure assessment. It is performed by combining the exposure and dose-response assessments. The summary effects of the uncertainties in the preceding steps are described in this step.

The relations among the four steps of risk assessment and between risk assessment and risk management are depicted in Figure I-1. The type of research information needed for each step is also illustrated.

SCIENTIFIC BASIS FOR RISK ASSESSMENT

Step 1. Hazard Identification

Although risk assessment as it is currently practiced by federal agencies for the estimation of carcinogenic risk contains several relatively new features, the scientific basis for much of the analysis done in risk assessment is well established. This is especially true of the first step in the assessment process, hazard identification. Four general classes of information may be used in this step: epidemiologic data, animal-bioassay data, data on in vitro effects, and comparisons of molecular structure.

Epidemiologic Data

Well-conducted epidemiologic studies that show a positive association between an agent and a disease are

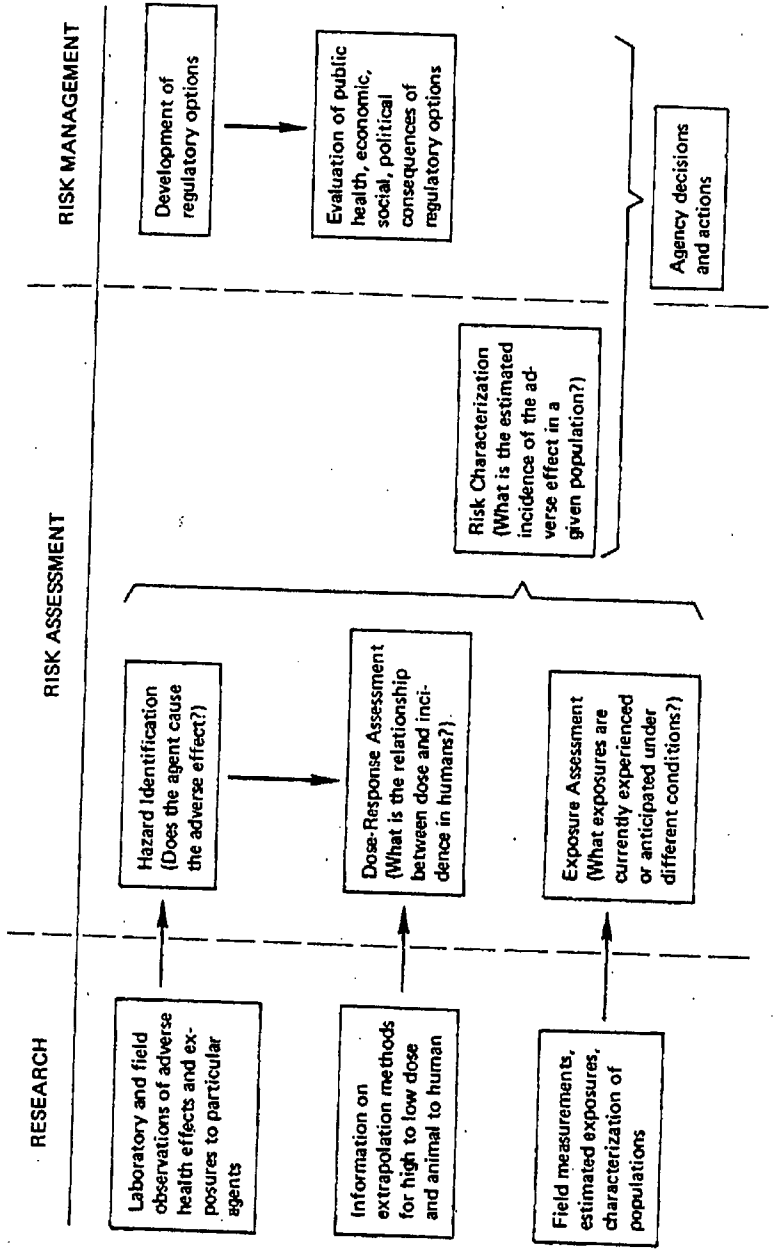

FIGURE I-1 Elements of risk assessment and risk management.

accepted as the most convincing evidence about human risk. This evidence is, however, difficult to accumulate; often the risk is low, the number of persons exposed is small, the latent period between exposure and disease is long, and exposures are mixed and multiple. Thus, epidemiologic data require careful interpretation. Even if these problems are solved satisfactorily, the preponderance of chemicals in the environment has not been studied with epidemiologic methods, and we would not wish to release newly produced substances only to discover years later that they were powerful carcinogenic agents. These limitations require reliance on less direct evidence that a health hazard exists.

Animal-Bioassay Data

The most commonly available data in hazard identification are those obtained from animal bioassays. The inference that results from animal experiments are applicable to humans is fundamental to toxicologic research; this premise underlies much of experimental biology and medicine and is logically extended to the experimental observation of carcinogenic effects. Despite the apparent validity of such inferences and their acceptability by most cancer researchers, there are no doubt occasions in which observations in animals may be of highly uncertain relevance to humans.

Consistently positive results in the two sexes and in several strains and species and higher incidences at higher doses constitute the best evidence of carcinogenicity. More often than not, however, such data are not available. Instead, because of the nature of the effect and the limits of detection of animal tests as they are usually conducted, experimental data leading to a positive finding sometimes barely exceed a statistical threshold and may involve tumor types of uncertain relation to human carcinogenesis. Interpretation of some animal data may therefore be difficult. Notwithstanding uncertainties associated with interpretation of some animal tests, they have, in general, proved to be reliable indicators of carcinogenic properties and will continue to play a pivotal role in efforts to identify carcinogens.

Short-Term Studies

Considerable experimental evidence supports the proposition that most chemical carcinogens are mutagens and that many mutagens are carcinogens. As a result, a positive response in a mutagenicity assay is supportive

evidence that the agent tested is likely to be carcinogenic. Such data, in the absence of a positive animal bioassay, are rarely, if ever, sufficient to support a conclusion that an agent is carcinogenic. Because short-term tests are rapid and inexpensive, they are valuable for screening chemicals for potential carcinogenicity and lending additional support to observations from animal and epidemiologic investigations.

Comparisons of Molecular Structure

Comparison of an agent's chemical or physical properties with those of known carcinogens provides some evidence of potential carcinogenicity. Experimental data support such associations for a few structural classes; however, such studies are best used to identify potential carcinogens for further investigation and may be useful in priority-setting for carcinogenicity testing.

Step 2. Dose-Response Assessment

In a small number of instances, epidemiologic data permit a dose-response relation to be developed directly from observations of exposure and health effects in humans. If epidemiologic data are available, extrapolations from the exposures observed in the study to lower exposures experienced by the general population are often necessary. Such extrapolations introduce uncertainty into the estimates of risk for the general population. Uncertainties also arise because the general population includes some people, such as children, who may be more susceptible than people in the sample from which the epidemiologic data were developed.

The absence of useful human data is common for most chemicals being assessed for carcinogenic effect, and dose-response assessment usually entails evaluating tests that were performed on rats or mice. The tests, however, typically have been designed for hazard identification, rather than for determining dose-response relations. Under current testing practice, one group of animals is given the highest dose that can be tolerated, a second group is exposed at half that dose, and a control group is not exposed. (The use of high doses is necessary to maximize the sensitivity of the study for determining whether the agent being tested has carcinogenic potential.) A finding in such studies that increased exposure leads to an increased incidence has been used primarily

to corroborate hazard identification, that is, to show that the agent does indeed induce the adverse health effect.

The testing of chemicals at high doses has been challenged by some scientists who argue that metabolism of chemicals differs at high and low doses; i.e., high doses may overwhelm normal detoxification mechanisms and provide results that would not occur at the lower doses to which humans are exposed. An additional factor that is often raised to challenge the validity of animal data to indicate effects in man is that metabolic differences among animal species should be considered when animal test results are analyzed. Metabolic differences can have important effects on the validity of extrapolating from animals to man if, for example, the actual carcinogen is a metabolite of the administered chemical and the animals tested differ markedly from humans in their production of that metabolite. A related point is that the actual dose of carcinogen reaching the affected tissue or organ is usually not known; thus, dose-response information, of necessity, is based on administered dose and not tissue dose. Although data of these types would certainly improve the basis for extrapolating from high to low doses and from one species to another, they are difficult to acquire and often unavailable.

Regulators are interested in doses to which humans might be exposed, and such doses usually are much lower than those administered in animal studies. Therefore, dose-response assessment often requires extrapolating an expected response curve over a wide range of doses from one or two actual data points. In addition, differences in size and metabolic rates between man and laboratory animals require that doses used experimentally be converted to reflect these differences.

Low-Dose Extrapolation

One may extrapolate to low doses by fitting a mathematical model to animal dose-response data and using the model to predict risks at lower doses corresponding to those experienced by humans. At present, the true shape of the dose-response curve at doses several orders of magnitude below the observation range cannot be determined experimentally. Even the largest study on record--the ED_{01} study involving 24,000 animals--was designed only to measure the dose corresponding to a 1% increase in tumor incidence. However, regulatory agencies are often concerned about much lower risks (1 in 100,000 to 1

in 1,000). Several methods have been developed to extrapolate from high doses to low doses that would correspond to risk of such magnitudes. A difficulty with low-dose extrapolation is that a number of the extrapolation methods fit the data from animal experiments reasonably well, and it is impossible to distinguish their validity on the basis of goodness of fit. (From a mathematical point of view, distinguishing among these models on the basis of their fit with experimental data would require an extremely large experiment; from a practical point of view, it is probably impossible). As Figure I-2 shows, the dose-response curves derived with different models diverge substantially in the dose range of interest to regulators. Thus, low-dose extrapolation must be more than a curve-fitting exercise, and considerations of biological plausibility must be taken into account.

Although the five models shown in Figure I-2 may fit experimental data equally well, they are not equally plausible biologically. Most persons in the field would agree that the supralinear model can be disregarded, because it is very difficult to conceive of a biologic mechanism that would give rise to this type of low-dose response. The threshold model is based on the assumption that below a particular dose (the "threshold" dose of a given carcinogen) there is no adverse effect. This concept is plausible but not now confirmable. The ED_{01} study showed an apparent threshold for bladder cancers caused by 2-acetylaminofluorene; when the data were replotted on a scale giving greater resolution (OTA, 1981), the number of bladder tumors consistently increased with dose, even at the lowest doses, and no threshold was detected. Another aspect of the debate over thresholds for inducing carcinogenic effects is the argument that agents that act through genotoxic mechanisms are not likely to have a threshold, whereas agents whose effects are mediated by epigenetic mechanisms are possibly more likely to have a threshold. The latter argument is also currently open to scientific challenge. Finally, apparent thresholds observable in animal bioassays cannot be equated with thresholds for entire populations. Even if a threshold exists for individuals, a single threshold would probably not be applicable to the whole population.

Animal-to-Human Dose Extrapolation

In extrapolating from animals to humans, the doses used in bioassays must be adjusted to allow for differ-

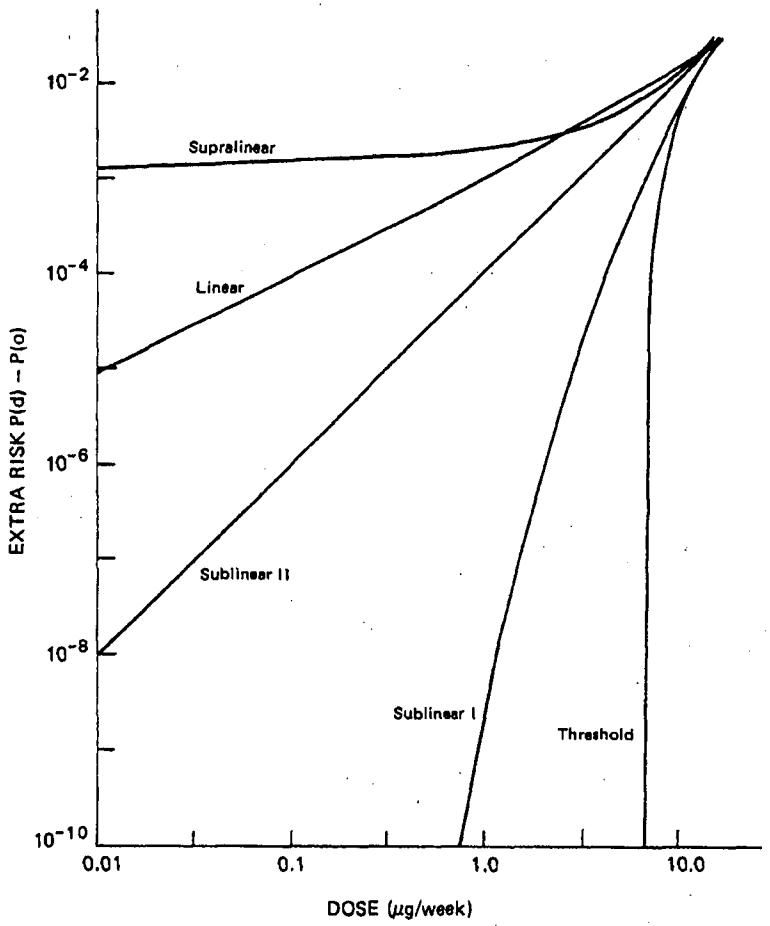

FIGURE I-2 Results of alternative extrapolation models for the same experimental data. NOTE: Dose-response functions were developed (Crump, in press) for data from a benzopyrene carcinogenesis experiment with mice conducted by Lee and O'Neill (1971).

ences in size and metabolic rates. Several methods currently are used for this adjustment and assume that animal and human risks are equivalent when doses are measured as milligrams per kilogram per day, as milligrams per square meter of body surface area, as parts per million in air, diet, or water, or as milligrams per kilogram per lifetime. Although some methods for conversion are used more frequently than others, a scientific basis for choosing one over the other is not established.

Step 3. Exposure Assessment

The first task of an exposure assessment is the determination of the concentration of the chemical to which humans are exposed. This may be known from direct measurement, but more typically exposure data are incomplete and must be estimated. Models for estimating exposure can be complex, even in the case of structured activity, as occurs in the workplace. Exposure measurements made on a small group (e.g., workers in a particular industrial firm) are often applied to other segments of the worker population.

Exposure assessment in an occupational setting consists primarily of estimation of long-term airborne exposures in the workplace. However, because an agent may be present at various concentrations in diverse occupational settings, a census of exposures is difficult and costly to conduct. In the community environment, the ambient concentrations of chemicals to which people may be exposed can be estimated from emission rates only if the transport and conversion processes are known. Alternative engineering control options require different estimates of the reduction in exposure that may be achieved. For new chemicals with no measurement data at all, rough estimations of exposure are necessary. Some chemical agents are of concern because they are present in foods or may be absorbed when a consumer product is used. Assessments of exposure to such agents are complicated by variations in diet and personal habits among different groups in the population. Even when the amount of an agent in a food can be measured, differences in food storage practices, food preparation, and dietary frequency often lead to a wide variation in the amount of the agent that individuals ingest. Patterns of use affect exposure to many consumer products; for example, a solvent whose vapor is potentially toxic may be used outdoors or it may be used in a small, poorly ventilated room, where the concentration of vapor in the air is much higher.

Another important aspect of exposure assessment is the determination of which groups in the population may be exposed to a chemical agent; some groups may be especially susceptible to adverse health effects. Pregnant women, very young and very old people, and persons with impaired health may be particularly important in exposure assessment. The importance of exposures to a mixture of carcinogens is another factor that needs to be considered in asssing human exposures. For example, exposure to cigarette smoke and asbestos gives an incidence of cancer that is much greater than anticipated from carcinogenicity data on each substance individually. Because data detecting such synergistic effects are often unavailable, they are often ignored or accounted for by the use of various safety factors.

Step 4. Risk Characterization

Risk characterization, the estimate of the magnitude of the public-health problem, involves no additional scientific knowledge or concepts. However, the exercise of judgment in the aggregation of population groups with varied sensitivity and different exposure may affect the estimate.

SCIENTIFIC AND POLICY JUDGMENTS IN RISK ASSESSMENT

The uncertainties inherent in risk assessment can be grouped in two general categories: missing or ambiguous information on a particular substance and gaps in current scientific theory. When scientific uncertainty is encountered in the risk assessment process, inferential bridges are needed to allow the process to continue. The Committee has defined the points in the risk assessment process where such inferences must be made as components. The judgments made by the scientist/risk assessor for each component of risk assessment often entail a choice among several scientifically plausible options; the Committee has designated these inference options.

COMPONENTS OF RISK ASSESSMENT

A list of components in carcinogenicity risk assessments was compiled by the Committee and is given below. This

list is not exhaustive or comprehensive, nor would all components listed be found in every risk assessment. The actual array of components in a particular risk assessment depends on a number of factors, including the types and extent of available data.

Hazard Identification

Epidemiologic Data

- What relative weights should be given to studies with differing results? For example, should positive results outweigh negative results if the studies that yield them are comparable? Should a study be weighted in accord with its statistical power?
- What relative weights should be given to results of different types of epidemiologic studies? For example, should the findings of a prospective study supersede those of a case-control study, or those of a case-control study those of an ecologic study?
- What statistical significance should be required for results to be considered positive?
- Does a study have special characteristics (such as the questionable appropriateness of the control group) that lead one to question the validity of its results?
- What is the significance of a positive finding in a study in which the route of exposure is different from that of a population at potential risk?
- Should evidence on different types of responses be weighted or combined (e.g., data on different tumor sites and data on benign versus malignant tumors)?

Animal-Bioassay Data

- What degree of confirmation of positive results should be necessary? Is a positive result from a single animal study sufficient, or should positive results from two or more animal studies be required? Should negative results be disregarded or given less weight?
- Should a study be weighted according to its quality and statistical power?
- How should evidence of different metabolic pathways or vastly different metabolic rates between animals and humans be factored into a risk assessment?
- How should the occurrence of rare tumors be treated? Should the appearance of rare tumors in a treated group be considered evidence of carcinogenicity even if the finding is not statistically significant?

- How should experimental-animal data be used when the exposure routes in experimental animals and humans are different?
- Should a dose-related increase in tumors be discounted when the tumors in question have high or extremely variable spontaneous rates?
- What statistical significance should be required for results to be considered positive?
- Does an experiment have special characteristics (e.g., the presence of carcinogenic contaminants in the test substance) that lead one to question the validity of its results?
- How should findings of tissue damage or other toxic effects be used in the interpretation of tumor data? Should evidence that tumors may have resulted from these effects be taken to mean that they would not be expected to occur at lower doses?
- Should benign and malignant lesions be counted equally?
- Into what categories should tumors be grouped for statistical purposes?
- Should only increases in the numbers of tumors be considered, or should a decrease in the latent period for tumor occurrence also be used as evidence of carcinogenicity?

Short-Term Test Data

- How much weight should be placed on the results of various short-term tests?
- What degree of confidence do short-term tests add to the results of animal bioassays in the evaluation of carcinogenic risks for humans?
- Should in vitro transformation tests be accorded more weight than bacterial mutagenicity tests in seeking evidence of a possible carcinogenic effect?
- What statistical significance should be required for results to be considered positive?
- How should different results of comparable tests be weighted? Should positive results be accorded greater weight than negative results?

Structural Similarity to Known Carcinogens

- What additional weight does structural similarity add to the results of animal bioassays in the evaluation of carcinogenic risks for humans?

General

- What is the overall weight of the evidence of carcinogenicity? (This determination must include a judgment of the quality of the data presented in the preceding sections.)

Dose-Response Assessment

Epidemiologic Data

- What dose-response models should be used to extrapolate from observed doses to relevant doses?
- Should dose-response relations be extrapolated according to best estimates or according to upper confidence limits?
- How should risk estimates be adjusted to account for a comparatively short follow-up period in an epidemiologic study?
- For what range of health effects should responses be tabulated? For example, should risk estimates be made only for specific types of cancer that are unequivocally related to exposure, or should they apply to all types of cancers?
- How should exposures to other carcinogens, such as cigarette smoke, be taken into consideration?
- How should one deal with different temporal exposure patterns in the study population and in the population for which risk estimates are required? For example, should one assume that lifetime risk is only a function of total dose, irrespective of whether the dose was received in early childhood or in old age? Should recent doses be weighted less than earlier doses?
- How should physiologic characteristics be factored into the dose-response relation? For example, is there something about the study group that distinguishes its response from that of the general population?

Animal-Bioassay Data

- What mathematical models should be used to extrapolate from experimental doses to human exposures?
- Should dose-response relations be extrapolated according to best estimates or according to upper confidence limits? If the latter, what confidence limits should be used?
- What factor should be used for interspecies conversion of dose from animals to humans?

- How should information on comparative metabolic processes and rates in experimental animals and humans be used?
- If data are available on more than one nonhuman species or genetic strain, how should they be used? Should only data on the most sensitive species or strain be used to derive a dose-response function, or should the data be combined? If data on different species and strains are to be combined, how should this be accomplished?
- How should data on different types of tumors in a single study be combined? Should the assessment be based on the tumor type that was affected the most (in some sense) by the exposure? Should data on all tumor types that exhibit a statistically significant dose-related increase be used? If so, how? What interpretation should be given to statistically significant decreases in tumor incidence at specific sites?

Exposure Assessment*

- How should one extrapolate exposure measurements from a small segment of a population to the entire population?
- How should one predict dispersion of air pollutants into the atmosphere due to convection, wind currents, etc., or predict seepage rates of toxic chemicals into soils and groundwater?
- How should dietary habits and other variations in lifestyle, hobbies, and other human activity patterns be taken into account?
- Should point estimates or a distribution be used?
- How should differences in timing, duration, and age at first exposure be estimated?
- What is the proper unit of dose?
- How should one estimate the size and nature of the populations likely to be exposed?
- How should exposures of special risk groups, such as pregnant women and young children, be estimated?

*Current methods and approaches to exposure assessment appear to be medium- or route-specific. In contrast with hazard identification and dose-response assessment, exposure assessment has very few components that could be applicable to all media.

Risk Characterization

- What are the statistical uncertainties in estimating the extent of health effects? How are these uncertainties to be computed and presented?
- What are the biologic uncertainties in estimating the extent of health effects? What is their origin? How will they be estimated? What effect do they have on quantitative estimates? How will the uncertainties be described to agency decision-makers?
- Which dose-response assessments and exposure assessments should be used?
- Which population groups should be the primary targets for protection, and which provide the most meaningful expression of the health risk?

THE INTERPLAY OF SCIENCE AND POLICY IN RISK ASSESSMENT

A key premise of the proponents of institutional separation of risk assessment is that removal of risk assessment from the regulatory agencies will result in a clear demarcation of the science and policy aspects of regulatory decision-making. However, policy considerations inevitably affect, and perhaps determine, some of the choices among the inference options. To examine the types of judgments required in risk assessment, the Committee has analyzed several components and the inference options for each.

Hazard Identification

The Committee has identified 25 components in hazard identification. These components differ in a number of ways. However, two major differences germane to the question considered here are the degree of scientific uncertainty encountered in each and the effect of choosing different inference options on the outcome of the risk assessment. Consider the following examples.

One component of risk assessment is the decision as to whether to use experimental animal data to infer risks to humans. Although data from studies of rats and mice may not always be predictive of adverse health effects in humans, the scientific validity of this approach is widely accepted. The use of positive animal data is the more conservative choice for this component. The use of

negative animal data to determine the absence of carcinogenic risk is less conservative, especially when the sensitivity of the assay is low. (The Committee uses the term conservative with appropriate modifiers to describe the degree to which a particular inference option for components in hazard identification will increase the likelihood that a substance will be judged to be a significant hazard to human health).

A component about which there is considerably more scientific uncertainty than the preceding example is the question of whether to count all types of benign tumors as evidence of carcinogenicity. Some benign tumors probably can progress to malignant lesions and some probably do not. The judgment that benign tumors and malignant tumors should be counted equally will affect tumor incidence and may influence the yes-no determination in hazard identification, and it can also affect the dose-response relation by increasing incidence at the doses tested. Thus, counting benign tumors is often the more conservative approach.

The examples just given differ in the degree to which scientific understanding can inform the judgments to be made. They are similar, however, in that for each, the available inference options differ in conservatism. For many components, this difference in degree of conservatism among plausible inference options is not as clear as in the preceding examples and depends on the data available on a given substance. For example, the decision to combine incidences for all tumor types and calculate an overall tumor incidence can influence the final yes-no decision in hazard identification. However, in this case, whether such a choice is more conservative than not combining incidences depends on the incidences for each tumor type in test and control animals. If the incidence in control animals is slightly below the incidences in test animals for all tumor types and individual differences are not statistically significant, combining all tumor types would be more conservative. However, if incidences show no consistent trend and differences are statistically significant for only one tumor type, combining the tumors would be less conservative.

Dose-Response Assessment

The Committee has identified 13 components of dose-response assessment. Two major components are high- to low-dose extrapolation and interspecies dose conversion.

In a recent NRC report on the health effects of nitrate, nitrite, and N-nitroso compounds (National Academy of Sciences, 1981), three extrapolation models (the one-hit model, the multistage model, and the multihit model) were used to estimate the dose of a carcinogenic nitrosamine (dimethylnitrosamine) needed to cause cancer in one of a million rats. The doses calculated were 0.03 parts per billion (one-hit), 0.04 ppb (multistage), and 2.7 ppb (multihit); that is, the risk estimate per unit of dose would be higher for the one-hit and multistage models than for the multihit model for this experiment.

Other judgments in dose-response assessment that will affect the final estimate include choice of the experimental data set (from among many that might be available) to be used to calculate the relation between dose and incidence of tumors (e.g., use of the most sensitive animal group will result in the most conservative estimate), choice of a scaling factor for conversion of doses in animals to humans (the risks calculated can vary by a factor of up to 35, depending on the method used), and the decision of whether to combine tumor types in determining incidence (as mentioned earlier, the decision to lump tumors might be more or less conservative than the decision not to combine incidences from different tumor types).

Exposure Assessment

Discussion of specific components in exposure assessment is complicated by the fact that current methods and approaches to exposure assessment appear to be medium- or route-specific. In contrast with hazard identification and dose-response assessment, exposure assessment has very few components that could be applicable to all media. For example, a model describing transport of a chemical through the atmosphere is necessarily quite different from a model describing transport through water or soil, whereas the use of a particular dose-response extrapolation model in dose-response assessment is independent of the medium or route of exposure. In any event, an assessor has several options available for estimating exposure to a particular agent in a particular medium, and these options will yield more or less conservative estimates of exposure. Among the options are different assumptions about the frequency and duration of human

exposure to an agent or medium, rates of intake or contact, and rates of absorption.

Risk Characterization

The final expressions of risk derived in this step will be used by the regulatory decision-maker when health risks are weighed against other societal costs and benefits to determine an appropriate action. Little guidance is available on how to express uncertainties in the underlying data and on which dose-response assessments and exposure assessments should be combined to give a final estimate of possible risk.

Basis for Selecting Inference Options

The Committee has presented some of the more familiar, and possibly more controversial, components of risk assessment. A review of the list of components reveals that many components lack definitive scientific answers, that the degree of scientific consensus concerning the best answer varies (some are more controversial among scientists than others), and that the inference options available for each component differ in their degree of conservatism. The choices encountered in risk assessment rest, to various degrees, on a mixture of scientific fact and consensus, on informed scientific judgment, and on policy determinations (the appropriate degree of conservatism).

That a scientist makes the choices does not render the judgments devoid of policy implications. Scientists differ in their opinions of the validity of various options, even if they are not consciously choosing to be more or less conservative. In considering whether to use data from the most sensitive experimental animals for risk assessment, a scientist may be influenced by the species, strains, and gender of the animals tested, the characteristics of the tumor, and the conditions of the experiment. A scientist's weighting of these variables may not easily be expressed explicitly, and the result is a mixture of fact, experience (often called intuition), and personal values that cannot be disentangled easily. As a result, the choice made may be perceived by the scientist as based primarily on informed scientific judgment. From a regulatory official's point of view, the same choice

may appear to be a value decision as to how conservative regulatory policy should be, given the lack of a decisive empirical basis for choice.

A risk assessor, in the absence of a clear indication based on science, could choose a particular approach (e.g., the use of an extrapolation model) solely on the basis of the degree to which it is conservative, i.e., on the basis of its policy implications. Furthermore, a desire to err on the side of overprotection of public health by increasing the estimate of risk could lead an assessor to choose the most conservative assumptions throughout the process for components on which science does not indicate a preferred choice. Such judgments made in risk assessment are designated <u>risk assessment policy</u>, that is, policy related to and subservient to the scientific content of the process, in contrast with policy invoked to guide risk management decisions, which has political, social, and economic determinants.

When inference options are chosen primarily on the basis of policy, risk management considerations (the desire to regulate or not to regulate) may influence the choices made by the assessors. The influence can be generic or ad hoc; i.e., assessments for <u>all</u> chemicals would consistently use the more or less conservative inference options, depending on the <u>overall</u> policy orientation of the agency ("generic"), or assessments would vary from chemical to chemical, with more conservative options being chosen for substances that the agency wishes to regulate and less conservative options being chosen for substances that the agency does not wish to regulate. (The desire to regulate or not would presumably stem from substance-specific economic and social considerations.) The possible influence of risk management considerations, whether real or perceived, on the policy choices made in risk assessment has led to reform proposals (reviewed later in this report) that would separate risk assessment activities from the regulatory agencies.

Table I-1 recapitulates the terms introduced in this discussion.

RISK ASSESSMENT IN PRACTICE

This section addresses past agency practices of risk assessment associated with efforts to regulate toxic substances.

TABLE I-1 Summary of Terms

Risk Assessment. Risk assessment is the qualitative or quantitative characterization of the potential health effects of particular substances on individuals or populations.
Risk Management. Risk management is the process of evaluating alternative regulatory options and selecting among them. A risk assessment may be one of the bases of risk management.
Steps. Risk assessments comprise many or all of the following steps: hazard identification, dose-response assessment, exposure assessment, and risk characterization.
Components. Steps in risk assessment comprise many components--points in a risk assessment at which judgments must be made regarding the analytic approach to be taken.
Inference options. For many components, two or more inference options are available.
Risk Assessment Policy. Risk assessment policy consists of the analytic choices that must be made in the course of a risk assessment. Such choices are based on both scientific and policy considerations.

RISK ASSESSMENT AND REGULATORY DECISION-MAKING

The regulatory process can be initiated in many ways. Each regulatory agency typically has jurisdiction over a large number of substances, but circumstances force an allocation of resources to a few at a time. The decision as to which substances to regulate is based, at least in part, on the degree of hazard. Thus, some notion of relative hazard (implicit or explicit, internally generated or imposed by outside groups) is necessary. Critics of federal regulation have contended that the agencies have not set their priorities sensibly. In general, agency risk assessments for priority-setting have been more informal, less systematic, and less visible than those for establishing regulatory controls.

Agenda-setting involves decisions about which substances should be selected (and often in what order) for more intense formal regulatory review. All programs face this problem, but it assumes different configurations: some programs cover a finite and known set of chemicals that must be reviewed, so the order of the regulatory reviews is the key question, and the primary job of the risk assessor is to help the agency implement a worst-first approach. For example, EPA's pesticides program has long had lists of suspect pesticide ingredients, and agency officials have had to decide which ones warrant formal consideration of cancellation or of new controls. An agency's agenda may also respond to private-sector initiatives (in the case of approval of new drugs or pesticides), conform to statutory directives, or react to new evidence of hazards previously unrecognized or thought to be less serious. This agenda formation phase, too, involves elements of risk assessment by the agency, the Congress, or private-sector entities; that is, there must be some assessment, however informal, that indicates reason for concern.

For many items on an agency's regulatory agenda, hazard identification alone will support a conclusion that a chemical presents little or no risk to human health and should be removed from regulatory consideration, at least until new data warrant renewed concern. If a chemical is found to be potentially dangerous in the hazard-identification step, it could then be taken through the steps of dose-response assessment, exposure assessment, and risk characterization. At any of these steps, the evaluation might indicate that a substance poses little or no risk and therefore can be removed from regulatory consideration until new data indicate a need for reevaluation.

Chemicals that are judged to present appreciable risks to health are candidates for regulatory action, and an agency will begin to develop options for regulating exposures. Regulatory options usually involve specific product or process changes and typically need to be based on extensive engineering and technical knowledge of the affected industry. Evaluation of the regulatory options includes recomputation of the predicted risk, in accord with altered expectations of exposure intensity or numbers of persons exposed.

Many of the activities of regulatory agencies do not conform to this sequential approach. However, regardless of the sequence of steps and the number of steps used to

determine whether regulatory action is warranted, risk assessment serves at least two major functions in regulatory decisions: first, it provides an initial assessment of risks, and, if the risk is judged to be important enough to warrant regulatory action, it is used to evaluate the effects of different regulatory options on exposure. In addition, it may be used to set priorities for regulatory consideration and for further toxicity testing.

These varied functions place different requirements on risk assessors, and a single risk assessment method may not be sufficient. A risk assessment to establish testing priorities may appropriately incorporate many worst-case assumptions if there are data gaps, because research should be directed at substances with the most crucial gaps; but such assumptions may be inappropriate for analyzing regulatory controls, particularly if the regulator must ensure that controls do not place undue strains on the economy. In establishing regulatory priorities, the same inference options should be chosen for all chemicals, because the main point of the analysis is to make useful risk comparisons so that agency resources will be used rationally. However, this approach, which may be reasonable for priority-setting, may have to yield to more sophisticated and detailed scientific arguments when a substance's commercial life is at stake and the agency's decision may be challenged in court. Furthermore, the available resources and the resulting analytic care devoted to a risk assessment for deciding regulatory policy are likely to be much greater for analyzing control actions for a single substance than for setting priorities.

THE AGENCIES THAT REGULATE

The approach to risk assessment varies considerably among the four federal agencies. Differences stem primarily from variations in agency structure and differences in statutory mandates and their interpretation.

Organizational Arrangements

The Food and Drug Administration (FDA) is a component of the Department of Health and Human Services, whose Secretary is the formal statutory delegate of the powers exercised by FDA. FDA is headed by a single official,

the Commissioner of Food and Drugs, who is appointed by and serves at the pleasure of the Secretary of the Department of Health and Human Services. It is organized in product-related bureaus, each of which employs its own scientists, technicians, compliance officers, and administrators. FDA has a long (75-year) and strong scientific tradition. According to a recent Office of Technology Assessment summary, FDA had taken or proposed action on 24 potential carcinogens by 1981.

Like FDA, the Environmental Protection Agency (EPA) is headed by a single official, but EPA's Administrator is appointed by the President subject to Senate confirmation. Also like FDA, EPA resembles a confederation of relatively discrete programs that are coordinated and overseen by a central management. The agency was established in 1970, but many of its programs (e.g., air and water pollution control and pesticide regulation) predate its formation and previously were housed in and administered by other departments. Other programs, such as those for toxic substances and hazardous waste, are rather new. EPA's research, policy evaluation, and, until recently, enforcement efforts were separated organizationally from the program offices that write regulations. EPA has had the widest experience with regulating carcinogens; as of 1981, it had acted on 56 chemicals in its clean-water program, 29 in its clean-air program, 18 in its pesticide program, and two in its drinking-water program.

The Occupational Safety and Health Administration (OSHA) is part of the Department of Labor. The agency's head is an Assistant Secretary of Labor, who requires Senate confirmation. Although FDA and EPA derive their scientific support largely from their own full-time employees, until the late 1970s OSHA relied on other agencies, primarily the National Institute of Occupational Safety and Health, an agency of the Department of Health and Human Services. This division reflects a conscious congressional choice in 1970 to place the health experts on whom OSHA was expected to rely in an outside environment believed more congenial to scientific inquiry and less vulnerable to political influence. As of 1981, 18 potential carcinogens had been acted on by OSHA.

The Consumer Product Safety Commission (CPSC) enforces five statutes, including the Consumer Product Safety Act and the Federal Hazardous Substances Act. Both empower CPSC to regulate unreasonable risks of injury from products used by consumers in the home, in schools, or in

recreation. The much smaller CPSC differs sharply from the other three agencies in two important respects: it does not have a single administrative head, but instead is governed by five Commissioners, who can make major regulatory decisions only by majority vote; and the Commissioners are appointed for fixed terms by the President with Senate confirmation. Before 1981, CPSC had acted on five potential carcinogens.

The four agencies have attempted to coordinate risk assessment activities in the past, most notably through the Interagency Regulatory Liaison Group (IRLG), which formed a work group on risk assessment to develop a guideline for assessing carcinogenic risks. Assisted by scientists from the National Cancer Institute and the National Institute for Environmental Health Sciences, it examined the various approaches used by the four agencies to evaluate evidence of carcinogenicity and to assess risk. The IRLG (1979a,b) then integrated and incorporated these evaluative procedures into a document, "Scientific Bases for Identification of Potential Carcinogens and Estimation of Risks," which described the basis for evaluation of carcinogenic hazards identified through epidemiologic and experimental studies and the methods used for quantitative estimation of carcinogenic risk.

Regulatory Statutes*

Examination of the statutes that the four agencies administer reveals important differences in the standards that govern their decisions. The Office of Technology Assessment has summarized (Table I-2) statutes that pertain to the regulation of carcinogenic chemicals. In particular, the statutes accord different weights to such criteria as risk, costs of control, and technical feasibility. In addition, different modes of regulation vary in their capacity to generate the scientific data necessary to perform comprehensive risk assessments.

Several laws require agencies to balance regulatory costs and benefits. Examples of balancing provisions are found in the Safe Drinking Water Act; the Federal Insecticide, Fungicide, and Rodenticide Act; the Toxic Substances

*This discussion draws heavily on the Office of Technology Assessment report, Technologies for Determining Cancer Risks from the Environment, 1981.

Control Act; and the section on fuel additives in the Clean Air Act. Under such provisions, a risk assessment can be used to express the nature and extent of public-health benefits to be attained through regulation.

Some regulatory programs involve the establishment of technology-based exposure controls. This approach is followed, for example, in portions of the clean-water program and the part of the hazardous-wastes program that deals with waste-incineration standards. In such programs, a risk assessment may be used to show the human exposure that corresponds to a specific degree of risk or to calculate the risk remaining after control technologies are put in place.

Some statutes mandate control techniques to reduce risks to zero whenever hazard is affirmed. Such techniques include outright bans of products, as envisioned in the Delaney clause in the Federal Food, Drug, and Cosmetic Act. In addition, if the concept of a threshold below which carcinogens pose no risk is not accepted, strict interpretations of ample margin of safety language in federal clean-air and clean-water legislation would require that exposures to carcinogenic pollutants be reduced to zero. The role of risk assessment in cases where mandatory control techniques must reduce risks to zero may be simply to affirm that a hazard exists.

The difference between programs that involve premarketing approval of substances and programs that operate through post hoc mechanisms, such as environmental emission limits, may have an important influence over the quality of risk assessments. The most important effect of this difference may lie in the fact that premarketing approval programs (such as those for pesticides, for new human drugs, and for new food additives) empower an agency to require the submission of sufficient data for a comprehensive risk assessment, whereas other programs tend to leave agencies to fend for themselves in the acquisition of necessary data.

There can be little question that differing statutory standards for decision affect the weight that agencies accord risk assessments. Like differences in the mode of regulation, they probably have affected the rigor and scope of many assessments. If risk is but one of several criteria that a regulator must consider or if data are expensive to obtain, it would not be surprising if an agency devoted less effort to risk assessment. However, the Committee has not discovered differences in existing statutes that should impede the adoption of uniform,

TABLE I-2 Public Laws Providing for the Regulation of Exposures to Carcinogens

Legislation (Agency)	Definition of toxics or hazards used for regulation of carcinogens	Degree of protection	Agents regulated as carcinogens (or proposed for regulation)	Basis of the legislation	Remarks
Federal Food, Drug and Cosmetic Act: (FDA)					
Food	Carcinogenicity for *additive* defined by Delaney Clause	No risk permitted, ban of additive	21 food additives and colors	Risk	
	Contaminants	"necessary for the protection of public health..." sec. 406 (346)	Three substances—aflatoxin, PCBs, nitrosamines	Balancing	
Drugs	Carcinogenicity is defined as a risk	Risks and benefits of drug are balanced.	Not determined	Balancing	
Cosmetics	"substance injurious under conditions of use prescribed."	Action taken on the basis that cosmetic is adulterated.	Not determined	Risk. No health claims are allowed for "cosmetics." If claims are made, cosmetic becomes a "drug."	
Occupational Safety and Health Act (OSHA)	Not defined in Act (but OSHA Generic Cancer Policy defines carcinogens on basis of animal test results or epidemiology.)	"adequately assures to the extent feasible that no employee will suffer material impairment of health or functional capacity..." sec. 6(b) (5)	20 substances	Technology (or balancing)	
Clean Air Act (EPA)					
Sec. 112 (stationary sources)	"an air pollutant...which ...may cause, or contribute to, an increase in mortality or an increase in serious irreversible, or incapacitating reversible, illness." sec. 112(a) (1)	"an ample margin of safety to protect the public health..." sec. 112(b) (1) (B)	Asbestos, beryllium, mercury, vinyl chloride, benzene, radionuclides, and arsenic (an additional 24 substances are being considered)	Risk	Basis of the Airborne Carcinogen Policy
Sec. 202 (vehicles)	"air pollutant from any ...new motor vehicles... or engine, which...cause, or contribute to, air pollution which may reasonably be anticipated to endanger public health or welfare." sec. 202A(a) (1)	"standards which reflect the greatest degree of emission reduction achieveable through...technology ...available..." sec. 202(b) (3)(a) (1)	Diesel particulates standard	Technology Sec. 202(b) (4) (B) includes a risk-risk test for deciding between pollutant that might result from control attempts.	Sec. 202(b) (4) (A) specifies that no pollution control device, system, or element shall be allowed if it presents an unreasonable risk to health, welfare or safety.
Sec. 211 (fuel additives)	Same as above (211(c) (1)).	Same as above (211(c) (2) (a)).	—	Balancing. Technology-based with consideration of costs, but health-based in requirement that standards provide ample margin of safety.	A cost-benefit comparison of competing control technologies is required.
Clean Water Act (EPA) Sec. 307	Toxic pollutants listed in Committee Report 95-30 of House Committee on Public Works and Transportation. List from consent decree between EDF, NRDC, Citizens for Better Environment and EPA.	Defined by applying BAT economically achieveable (sec. 307(a) (2)), but effluent levels are to "provide(s) an ample margin of safety." (sec. 307(a) (4))	49 substances listed as carcinogens by CAG.	Technology	
Federal Insecticide, Fungicide, and Rodenticide Act and the Federal Environmental Pesticide Control Act (EPA)	One which results in "unreasonable adverse effects on the environment or will involve unreasonable hazard to the survival of a species declared endangered..."	Not specified.	14 rebuttable presumptions against registrations either initiated or completed; nine pesticides voluntarily withdrawn from market.	Sec. 2(bb) Balancing: "unreasonable adverse effects..."	"Unreasonable adverse effects" means "unreasonable risk to man or the environment taking into account the economic, social, and environmental costs and benefits..."

TABLE I-2 (Continued)

Legislation (Agency)	Definition of toxics or hazards used for regulation of carcinogens	Degree of protection	Agents regulated as carcinogens (or proposed for regulation)	Basis of the legislation	Remarks
Resource Conservation and Recovery Act (EPA)	One which "may cause, or significantly contribute to an increase in mortality or an increase in serious irreversible, or incapacitating reversible, illness; or, pose a...hazard to human health or the environment..." sec. 1004(5) (A) (B)	"that necessary to protect human health and the environment..." sec. 3002-04	74 substances proposed for listing as hazardous wastes	Risk. The Administrator can order monitoring and set standards for sites.	
Safe Drinking Water Act (EPA)	"contaminant(s) which...may have an adverse effect on the health of persons." sec. 1401(1) (B)	"to the extent feasible...(taking costs into consideration)..." sec. 1412(a) (2)	Trihalomethanes, chemicals formed by reactions between chlorine used as disinfectant and organic chemicals. Two pesticides and 2 metals classified as carcinogens by CAG, but regulated because of other toxicities.	Balancing	
Toxic Substances Control Act (EPA)					
Sec. 4 (to require testing)	substances which "may present an unreasonable risk of injury to health or the environment." sec. 4(a) (1) (A) (i)	Not specified.	Six chemicals used to make plastics pliable.	Balancing: "unreasonable risk"	
Sec. 6 (to regulate)	substances which "present(s) or will present an unreasonable risk of injury to health or the environment." sec. 6(a)	"to protect adequately against such risk using the least burdensome requirement" sec. 6(a)	PCBs regulated as directed by the law.	Balancing: "unreasonable risk."	
Sec. 7 (to commence civil action against imminent hazards)	"imminently hazardous chemical substance or mixture means a... substance or mixture which presents an imminent and unreasonable risk of serious or widespread injury to health or the environment."	Based on degree of protection in sec. 6			
Federal Hazardous Substances Act (CPSC)	"any substance (other than a radioactive substance) which has the capacity to produce personal injury or illness..." 15 USC sec.	"establish such reasonable variations or additional label requirements... necessary for the protection of public health and safety..." 15 USC sec.		Risk	"Highly toxic" defined as capacity to cause death, thus toxicity may be limited to acute toxicity.
Consumer Product Safety Act (CPSC)	"products which present unreasonable risks of injury...in commerce," and " 'risk of injury' means a risk of death, personal injury or serious or frequent injury." 15 USC sec. 2051 "imminently hazardous consumer product' means consumer product which presents imminent and unreasonable risk of death, serious illness or severe personal injury." 15 USC sec. 2061	"standard shall be reasonably necessary to prevent or reduce an unreasonable risk of injury." 15 USC sec. 2056	Five substances: asbestos, benzene, benzidine (and benzidine-based dyes and pigments), vinyl chloride, "tris"	Balancing: "unreasonable"	Standards are to be expressed, wherever feasible, as performance requirements.

SOURCE: Office of Technology Assessment, Technologies for Determining Cancer Risks from the Environment, 1981.

government-wide risk assessment guidelines. Indeed, it is not satisfied that there are legal bases for inter-agency differences in the performance--as distinct from the use--of risk assessment for chronic health hazards.

CONCLUSIONS

On the basis of a review of the nature and the policy context of risk assessment, the Committee has drawn the following general conclusions:

1. Risk assessment is only one aspect of the process of regulatory control of hazardous substances. Therefore, improvements in risk assessment methods cannot be assumed to eliminate controversy over federal risk management decisions.

Restrictive regulation has seemed onerous to manufacturers, distributors, and users of products judged useful and valuable; conversely, inaction and delay with respect to regulatory proceedings have appeared callous and irresponsible to others. These dissatisfactions have been manifested in many ways, including criticism of risk assessment processes. The Committee believes that much of this criticism is inappropriately directed and gives rise to an unrealistic expectation that modifying risk assessment procedures will result in regulatory decisions more acceptable to the critics. Certainly risk assessment can and should be improved, with salutary effects on the appropriateness of regulatory decisions. However, risk management, although it uses risk assessment, is driven by political, social, and economic forces, and regulatory decisions will continue to arouse controversy and conflict.

2. Risk assessment is an analytic process that is firmly based on scientific considerations, but it also requires judgments to be made when the available information is incomplete. These judgments inevitably draw on both scientific and policy considerations.

The primary problem with risk assessment is that the information on which decisions must be based is usually inadequate. Because the decisions cannot wait, the gaps in information must be bridged by inference and belief, and these cannot be evaluated in the same way as facts. Improving the quality and comprehensiveness of knowledge is by far the most effective way to improve risk assessment, but some limitations are inherent and unresolvable, and inferences will always be required. Although we conclude that the mixing of science and policy in risk assessment cannot be eliminated, we believe that most of the intrusions of policy can be identified and that a major contribution to the integrity of the risk assessment process would be the development of a procedure to ensure that the judgments made in risk assessments, and the underlying rationale for such judgments, are made explicit.

3. Two kinds of policy can potentially affect risk assessment: that which is inherent in the assessment process itself and that which governs the selection of regulatory options. The latter, risk management policy, should not be allowed to control the former, risk assessment policy.

Risk management policy, by its very nature, must entail value judgments related to public perceptions of risk and to information on risks, benefits, and costs of control strategies for each substance considered for regulation. Such information varies from substance to substance, so the judgments made in risk management must be case-specific. If such case-specific considerations as a substance's economic importance, which are appropriate to risk management, influence the judgments made in the risk assessment process, the integrity of the risk assessment process will be seriously undermined. Even the perception that risk management considerations are influencing the conduct of risk assessment in an important way will cause the assessment and regulatory decisions based on them to lack credibility.

4. Risk assessment suffers from the current absence of a mechanism for addressing generic issues in isolation from specific risk management decisions.

Although the practice of risk assessment has progressed in recent years, there is currently no mechanism for stimulating and monitoring advances on generic questions in relevant scientific fields or for the timely dissemination of such information to risk assessors.

REFERENCES

Crump, K. S. In press. Issues related to carcinogenic risk assessment from animal bioassay data. Paper

presented May 1981 at the International School of Technological Risk Assessment, a NATO Advanced Study Institute, Erice, Italy.

IRLG (Interagency Regulatory Liaison Group), Work Group on Risk Assessment. 1979a. Scientific bases for identification of potential carcinogens and estimation of risks. Fed. Reg. 44:39858.

IRLG (Interagency Regulatory Liaison Group), Work Group on Risk Assessment. 1979b. Scientific bases for identification of potential carcinogens and estimation of risks. J. Natl. Cancer Inst. 63:242.

Lee, P. N., and J. A. O'Neill. 1971. The effect both of time and dose applied on tumor incidence rate in benzopyrene skin painting experiments. Brit. J. Cancer 25:759-770.

National Academy of Sciences. 1981. The Health Effects of Nitrate, Nitrite, and N-Nitroso Compounds. Washington, D.C.: National Academy Press. 544 pp.

OTA (Office of Technology Assessment). 1981. Assessment of the Technologies for Determining Cancer Risks from the Environment. 240 pp.

II
Inference Guidelines for Risk Assessment

INTRODUCTION AND DEFINITIONS

An inference guideline* is an explicit statement of a predetermined choice among the options that arise in inferring human risk from data that are not fully adequate or not drawn directly from human experience. A guideline might, for example, specify the mathematical model to be used to estimate the effects of exposure at low doses from observations based on higher doses. The most important feature of guideline use is that it changes the risk assessment process from one in which inference options are selected on a substance-by-substance basis to one in which they are selected once and thereafter

*The Committee hopes to avoid any misunderstanding resulting from its use of the terms inference guideline and guideline (used for brevity in lieu of inference guideline). This terminology is potentially confusing, because guidelines can be understood as codified principles addressed to a particular subject matter, risk assessment, or as describing the legal weight of any codified standards or principles. For the Committee, it has the former meaning. Inference guidelines are the principles followed by risk assessors in interpreting and reaching judgments based on scientific data. (Thus, our inference guidelines are distinct from the standards for toxicologic and other testing standards that many regulatory agencies and scientific bodies have adopted to govern, or at least influence, the generation of data later used in risk assessment.)

For many lawyers, the term guideline connotes the weight to be given to any set of codified principles, not

applied to an entire series of chemicals. In the absence of guidelines, assessors may well select the same inference options for substance after substance in a given agency program, and a common set of inference options may emerge, in common law fashion, from their consistent application in the program. But even the continued use of the same set of inference options over time does not necessarily imply that the assessors would make the same choices for every substance. Furthermore, outsiders would have no way of knowing what the common set is. In contrast, the use of guidelines makes more evident the generic choice of inference options, which we have seen in Chapter I, is based on both scientific and risk assessment policy considerations.

HISTORY OF THE USE OF GUIDELINES

SAFETY EVALUATION GUIDELINES FOR EFFECTS OTHER THAN CANCER

The development and use of guidelines by a regulatory agency first became of major importance after Congress

---

only those addressed to risk assessment, in legal proceedings. The Food and Drug Administration, for example, has defined a guideline as an official pronouncement of the agency describing a procedure that satisfies legal requirements, but is not mandated by law. A more complete treatment of the distinction between binding regulations and other formal agency pronouncements appears in the section of this chapter entitled "Degree to Which Guidelines May Be Binding on an Agency and a Regulated Party."

The Committee has used the term guideline to describe the principles by which risk assessments are to be performed, because that is the term Congress used in the legislation that authorized this study. The Committee was asked to consider the feasibility of establishing uniform "risk assessment guidelines." There is no evidence that Congress was aware of the different meanings of the term. It obviously was seeking advice about the intellectual and scientific bases for codified principles for risk assessment, not the appropriate legal form for their adoption. Faced with possible confusion no matter which terminology it chose, the Committee has retained the language that Congress itself used to describe our inquiry.

enacted amendments to the Federal Food, Drug, and Cosmetics Act in the 1950s and early 1960s. These laws, as applied to noncarcinogenic agents, required that food additives, color additives, drugs for animals, and pesticides be shown to be safe under their intended conditions of use before premarket approval by the Food and Drug Administration (FDA). The agency developed guidelines to provide a systematic way to deal with the legal requirements embodied in the amendments. Although guidelines for the conduct of various types of toxicity tests received greatest notice, some attention was given to the problem of data interpretation for inferring human risk. For example, a 1959 publication written by several members of the FDA Division of Pharmacology, Appraisal of the Safety of Chemicals in Foods, Drugs, and Cosmetics, is devoted primarily to toxicity testing methods, but contains one chapter called "Some Interpretative Problems in Evaluating the Safety of Food Additives" (Lehman et al., 1959). Although that publication, which served as a guide for both FDA and the regulated industry for at least a decade, was never published as a regulation, it was widely accepted by affected industrial concerns.

In all cases except that of carcinogens, establishment of acceptable intakes was accomplished by applying safety factors to experimentally derived no-observed-effect exposures. Testing involved mostly the use of acute and subchronic (90-day) animal tests, although some long-term tests were required. The use of safety factors to establish acceptable intakes was also recommended by the Food Protection Committee of the National Research Council (NRC/NAS, 1970) and adopted by the Joint Food and Agricultural Organization and World Health Organization Expert Committees on Food Additives (1972) and Pesticide Residues (1965). This approach continues to be used for noncarcinogenic food additives and pesticides and, in slightly modified form, to define acceptable exposures to occupational and various environmental pollutants.

These methods of assigning acceptable limits of exposure imply that the application of safety factors of various magnitudes to experimentally derived no-observed-effect exposures will ensure low risk. The acceptable exposure, whether expressed as an acceptable daily intake for a food additive or as a permissible exposure limit for an occupational agent, is derived by imposing untested assumptions (e.g., about the likely nature of dose-response relations at low doses) and by drawing inferences from sparse data. Safety evaluation schemes may therefore

be classified as a set of guidelines that emphasize testing methods heavily and that afford methods of inference only scant attention.

Recent efforts have dealt more directly with developing guidelines for risk assessment of noncarcinogenic effects. The Environmental Protection Agency (EPA) has proposed guidelines for chemical mutagenesis (EPA, 1980a) and has collected public comments on them, but has yet to publish a final rule. In addition, the agency cosponsored two conferences with Oak Ridge National Laboratory on risk assessment methods for reproductive and teratogenic effects; the proceedings of the conferences have been published (ORNL/EPA, 1982). The Interagency Regulatory Liaison Group began to develop guidelines for risk assessment of reproductive and teratogenic effects, but the effort ceased with the disbanding of the group in 1981. The March of Dimes Birth Defects Foundation (1981) has published the proceedings of a conference dealing with guidelines for studies of human populations exposed to mutagenic and reproductive hazards. Despite the increasing interest in noncarcinogenic effects, methods of estimating the risk of these effects have not been the subject of major public and scientific debate; attention has been devoted mainly to carcinogenic risk assessment. Much more critical review of the inferential methods for assigning risks to noncarcinogenic agents is warranted.

GUIDELINES FOR CARCINOGENIC RISK

Until the late 1950s, few agents, either chemical or physical, had been regulated in this country on the basis of their carcinogenic action. One important regulated agent was ionizing radiation. Permissible exposures to radiation were set in a manner similar to that for noncarcinogenic agents, by application of safety factors applied to specified exposures. In the debate over health effects of radioactive fallout from atomic weapons tests in the 1950s, evidence to support a nonthreshold theory for cancer induction emerged. Evidence was also accumulated to indicate that the nonthreshold theory might be applicable to chemical carcinogens. It was in this context that Congress enacted statutes* in the 1950s and early

*The enactment of these statutes did not necessarily bring a unique new concept to FDA. In the early 1950s, before their enactment, the agency had prohibited three food additives on the grounds that they were found to be carcinogenic in test animals.

1960s that required FDA to ban the use of food and color additives shown to be carcinogenic. The assumption, which differed from that underlying safety evaluation of noncarcinogens, was that no exposure could be presumed safe. Thus, a full risk assessment scheme was not needed for carcinogens. The process stopped at hazard identification.

Many factors contributed to the later use of dose-response assessment, exposure assessment, and risk characterization to determine quantitative estimates of risk. One of these may have been the growing perception during the 1960s and 1970s that many kinds of risk could not be eliminated completely without unacceptable social and economic consequences. New laws reflecting this belief were enacted, and some agencies were required to balance the risk posed by carcinogenic agents against their perceived benefits. Quantitative risk assessment was the system developed to estimate the risk side of the balance. Over a period of 2 decades, various expert committees sponsored by government agencies and other organizations published numerous reports dealing with carcinogenicity evaluation. Most of these were state-of-the art reports on aspects of carcinogenicity inference, and many suggested guidelines for hazard identification. More recent reports have dealt explicitly with quantitative risk assessment. The impetus for producing these reports was probably a belief in the federal research and regulatory communities that some scientific principles related to carcinogenicity data evaluation had to be continually reexamined and reaffirmed. This belief pervaded the public-health establishment not only in the United States, but also in other countries and in the United Nations.

The following discussion examines efforts to develop and apply guidelines for the evaluation of carcinogenicity data by the federal regulatory agencies and the International Agency for Research on Cancer over the last decade--efforts that developed out of 2 decades of scientific consensus-building.

International Agency for Research on Cancer (IARC)

In 1971, the International Agency for Research on Cancer (IARC), an agency of the World Health Organization, began publication of a series of monographs on known and suspected carcinogens. These monographs are prepared by international groups of experts assembled by IARC, who critically review pertinent literature and draw conclusions regarding the carcinogenicity of various substances. The results of IARC reviews and evaluations are widely accepted. The guidelines used for evaluation by the IARC expert committees are set forth in the monographs. They are expressed in very general terms and are related to only six components of hazard identification, completely covered in six pages of text. A major feature of the guidelines is the presentation of criteria that classify the evidence of suspected carcinogens as sufficient or limited. The IARC allows the expert committees considerable latitude to evaluate many inference options on a case-by-case basis, although the agency appears to insist on adherence to the few stated guidelines.

Food and Drug Administration

The 1958 Food Additives Amendment to the Food, Drug, and Cosmetics Act prohibited the use of food additives found to be carcinogenic. The law was also interpreted as prohibiting FDA approval of any drug, for use in animals produced for human food, that had been shown to cause cancer. In 1962, by congressional amendment, FDA was permitted to approve the use of a carcinogenic animal drug if the agency was convinced that no residue of a drug would be found in edible tissues of the treated animals. Congress specified that FDA was to prescribe the analytic methods for verifying the absence of residues. This directive proved to be unworkable, for two reasons: progress in analytic chemistry was so rapid that approved methods of analysis quickly became obsolete and improved detection methods showed that no drug administered to animals is ever entirely absent from animal tissues. The problem of enforcing the 1962 amendment was highlighted in the early 1970s, when diethylstilbestrol residues were discovered in beef liver with highly sensitive, but as yet unapproved, analytic methods.

In an attempt to provide a consistent and predictable procedure for approving methods to search for drug residues, FDA proposed sensitivity-of-method guidelines in the form of regulations (FDA, 1973, 1977, 1979b). Rather than gear criteria to an analytic technique, the agency defined its standards in terms of risk. It proposed that any assay approved for controlling a carcinogenic drug must be capable of measuring residues that present more than an insignificant risk of cancer, and specified a 10^{-6} lifetime risk of cancer as a quantitative criterion of insignificance. If a drug sponsor could provide a detection method capable of measuring residues posing a risk of this magnitude or greater, FDA would ignore residues that could not be detected with this method. Thus, FDA found guidelines for quantitative estimation of risk necessary. FDA's sensitivity-of-method guidelines are unique in several ways. They address a narrow though complex set of issues encountered in regulating a single class of products, animal drugs. Although they deal to a large extent with testing, they were the first to address quantitative risk assessment methods, listing assumptions for dose-response assessment, exposure assessment, and risk characterization. And they are the only guidelines that attempt to establish a definition of significant risk. The guidelines have yet to be adopted, a decade after they were first proposed, but the agency has applied the methods of quantitative risk assessment embodied in the sensitivity-of-method document in connection with the regulation not only of animal drugs, but also of food contaminants, such as aflatoxin (FDA, 1979a) and trace constituents of some additives (FDA, 1982b). The sensitivity-of-method guidelines were proposed as regulations, as were the cancer guidelines of the Occupational Safety and Health Administration (OSHA). In both cases, regulation engendered substantial controversy. The major debate over the sensitivity-of-method guidelines has dealt not so much with risk assessment or the definition of significant risk as with the amount and cost of testing that FDA would require from industry before product approval.

Environmental Protection Agency

During the early to middle 1970s, EPA initiated actions to prohibit or restrict the use of several pesticides. The agency lacked internal procedures for assessing carcinogenic risk and relied heavily on the judgment of scientists outside EPA. Attorneys for EPA, in summar-

izing the testimony of their expert witnesses during administrative hearings on actions against the pesticides, set forth several statements that, in the legal brief, were referred to as cancer principles (EPA, 1972, 1975). They conveyed the idea that the only acceptable degree of regulation would be a total ban on exposures. The principles, perceived as EPA's cancer policy, incurred wide criticism from the scientific community, the private sector, and Congress. The impracticability of achieving zero risk on a broad scale for a large number of economically important chemicals became increasingly apparent. In response to this new perception, and perhaps out of a desire to avoid misunderstanding of its cancer policy, the EPA became the first agency to adopt formal guidelines embracing a two-step process of risk assessment (EPA, 1976). The first step is a determination of whether a particular substance constitutes a cancer risk (hazard identification). The second step is a determination of what regulatory action, if any, should be taken to reduce the risk. As part of the second step, the agency explicitly endorses the use of quantitative risk assessment as the means of determining the magnitude of the likely impact of a potential human carcinogen on public health. These guidelines were not published as regulations and enjoy fairly wide acceptance from most interested parties. As stated in the preface to the guidelines, they were published to improve agency procedures, to provide public notice of the approach that EPA would take, and to stimulate commentary from all sources on that approach. The guidelines were probably more important as a statement of a novel approach to risk assessment than for their content. They are quite general, cover less than a page of Federal Register text, and address only a few components of hazard identification, dose-response assessment, exposure assessment, and risk characterization. More detailed guidelines that specify assumptions for the choice of extrapolation models, scaling factors, and other elements of dose-response assessment were published in 1980 by program offices in EPA (EPA, 1980b). These rely in part on the Interagency Regulatory Liaison Group (IRLG) guidelines (IRLG, 1979a) and are currently undergoing review.

Occupational Safety and Health Administration

In 1977, OSHA published guidelines in a proposed regulation, "Identification, Classification, and Regulation of

Toxic Substances Posing a Potential Occupational Risk of Cancer" (OSHA, 1977); after extensive administrative hearings, it published a final rule in 1980 (OSHA, 1980). The guidelines proved to be highly controversial, and the hearings were marked by vigorous debate on almost every component of risk assessment covered by the guidelines.

The OSHA rule, written by agency staff, was a detailed scientific and regulatory document that took several hundred pages of Federal Register text and addressed almost every component of hazard identification. The final rule did not address exposure assessment and rejected the use of dose-response assessment for any regulatory purpose except priority-setting. The main purposes of the guidelines, as stated in the preface, were to streamline the process of risk assessment, to speed up regulation, and to reduce the workload of agency staff. Another purpose was to foster continuity of approach, even in the face of changes of policy-makers. OSHA staff perceived that the case-by-case approach to risk assessment was long and time-consuming, because the same controversial questions had to be argued each time a chemical was under consideration for regulation. The agency believed that the generic approach to risk assessment would reduce debate on these questions; the controversial issues could be decided once, incorporated into guidelines, and applied to all chemicals. For reasons of efficiency, the guidelines were written in language that permitted little deviation from the judgments embodied in them. Because they were written as regulations, regulated parties were required to abide by them. The agency has not used the rule as a basis for any published scientific assessment of carcinogenic hazard. It was revised in 1981 (OSHA, 1981) to accommodate the Supreme Court's ruling on benzene, which required that OSHA make a finding that the risk to workers in the absence of regulation was significant and would be reduced by the proposed standard. But this change and additional amendments were recently withdrawn, and the entire policy is under reconsideration (OSHA, 1982).

Consumer Product Safety Commission

The Consumer Product Safety Commission (CPSC) proposed cancer guidelines dealing mainly with hazard identification (CPSC, 1978). Ten components related to that step were addressed in several pages of Federal Register text.

Some minor attention was given to exposure assessment and dose-response assessment, for priority-setting purposes only. The rationale for publishing the guidelines, as stated in the preface of that document, was to establish CPSC's general principles and to solicit comments on them, to assist the general public and the regulated industries in understanding standards that CPSC would apply and regulatory actions that it was likely to take, and to set forth its approach to some issues that tended to recur in each case. The guidelines had no regulatory status; they were a statement of selected inference options to which the agency would adhere. The CPSC guidelines were never used; they were challenged in court, and the court ruled that CPSC had promulgated them illegally inasmuch as they were adopted without an opportunity for public comment. Furthermore, at that time CPSC had decided to rely on the guidelines of IRLG.

Interagency Regulatory Liaison Group

The four agencies represented in IRLG undertook the task of developing guidelines to "ensure that the regulatory agencies evaluate carcinogenic risk consistently." In 1979, after an 18-month interagency effort, IRLG published a report, "Scientific Bases for Identification of Potential Carcinogens and Estimation of Risk." The report was prepared by personnel of CPSC, EPA, FDA, and OSHA, with the assistance of senior scientists from the National Cancer Institute and the National Institute of Environmental Health Sciences. It was published in a scientific journal (IRLG, 1979b) and in the Federal Register (IRLG, 1979a); IRLG requested public comment in the Federal Register. The IRLG report was said to represent an interagency consensus on the scientific aspects of carcinogenic risk assessment.* It was the most comprehensive set of guidelines that had been developed for agency use, addressing most components of hazard identification and dose-response assessment with some general discussion of

*Because rule-making was under way in connection with its cancer policy, OSHA declined to participate in the IRLG notice and comment procedure. After the report was completed, the Food Safety Quality Service of the U.S. Department of Agriculture joined IRLG and participated in the notice and comment.

exposure assessment and risk characterization; it had, however, no official legal status. The report was noteworthy, in that it constituted the first evidence that all the federal regulatory agencies agreed on the inference options applicable to the identification of carcinogenic hazards and measurement of risks. The document made clear, however, that not all the agencies were bound to conduct quantitative assessments; it stated only that, if such assessments were to be conducted, they would be conducted uniformly. This language was probably a concession to OSHA's view, as expressed in its cancer policy, that quantitative risk assessment was to play no more than a priority-setting role in that agency's regulatory activities. Almost immediately after its publication, the IRLG report was adopted by the President's Regulatory Council and incorporated as the scientific basis of the Council's government-wide statement on regulation of chemical carcinogens. The Council viewed the IRLG guidelines as a major step in reducing inconsistency, duplication of effort, and lack of coordination among agencies in carcinogenic risk assessment (Regulatory Council, 1979).

The scientific aspects of the final OSHA cancer policy, which was written to allow less latitude in the choice of inference options, were, nevertheless, in general agreement with the IRLG guidelines. CPSC and EPA stated that they relied on the IRLG document, but the degree to which they rely on it today is not clear. FDA has made no statement other than that associated with the document's initial publication; in fact, in a recent proposal concerning the application of risk assessment to a class of trace constituents of additives, FDA did not even cite the IRLG document as a reference (FDA, 1982b). Although IRLG received a great deal of public comment on the guidelines, no report of the agencies' review of these comments has appeared. In fact, the document was heavily criticized by industry, because it was published in its final form and adopted before the comments could be analyzed and revisions incorporated. The Reagan Administration has officially disbanded the entire IRLG effort, so it is unlikely that review of the public comments will ever occur.

Although the IRLG charter was not renewed, a similar group has been established, but one that is coordinated by the White House Office of Science and Technology Policy. This group has prepared a draft document on the scientific basis of risk assessment and has distributed

it for comment (OSTP, 1982). The group anticipates that this document may serve as a reference point for later development of general guidelines for the agencies.

VARIATION IN THE FORM OF GUIDELINES

COMPREHENSIVENESS

Guidelines developed by agencies in the past have varied in the extent to which they have addressed each of the steps of risk assessment. IARC's guidelines address only hazard identification; OSHA's guidelines (1980) dealt mainly with hazard identification, with some discussion of dose-response assessment and none of exposure assessment and risk characterization; and IRLG's guidelines focused in detail on hazard identification and dose-response assessment, with some discussion of exposure assessment and risk characterization.

Guidelines also have varied in the extent to which they have addressed the components of the risk assessment steps. IARC's guidelines address a small number of components. Study of the latest IARC monograph (1982) reveals only six selected options that deal with inference of risk: treatment of benign versus malignant tumors, the choice of statistical methods for application of data from animal studies, the relevance of negative results of epidemiologic studies, the evaluation of tumors that occur spontaneously, the utility of short-term tests, and the overall weighting of evidence. The OSHA (1980) and IRLG documents, in contrast, each discussed and embraced over 20 selected options dealing with hazard identification.

EXTENT OF DETAIL

Guidelines have differed not only in their comprehensiveness, but also in the detail with which they have treated specific components of risk assessment. When the content of a guideline is detailed, the assessor is presented with more complete information than would be available from a more general guideline. For example, the statement in IARC's guidelines on benign tumors is general, compared with that in the IRLG guidelines. IARC concludes briefly:

> If a substance is found to induce only benign tumours in experimental animals, it should be

> suspected of being a carcinogen and requires further investigation.

The IRLG document made a similar statement, but in addition elaborated on several issues relevant to the evaluation of benign tumors that are not mentioned by IARC--e.g., evaluation of tumor incidence when both benign and malignant tumors are present; a listing of tumor types commonly observed as benign in test animals, but known to progress to frank malignant stages; evaluation of the quality of a histologic examination that might be presented as evidence; and an illustrative example of the dependence of response on the genetic characteristics of the test animal. The additional material could have been used by an assessor, particularly one not familiar with the newest information on benign tumors, to ensure that a more thorough analysis of the relevant issues had been performed.

FLEXIBILITY

Detail can often be confused with inflexibility, and it is important to make a distinction between these characteristics. Certainly, detailed guidelines can be inflexible if the detail is designed to limit agency discretion, and thus public debate, on an issue that is subject to multiple scientific interpretations. However, detailed guidelines can have quite a different effect if their intent is to provide an assessor with background information that describes the complexity of an issue, with nuances that may influence particular judgments, or with examples of cases that are legitimate exceptions to the general rule.

As described in Chapter I, almost all components of risk assessment theoretically embrace one or more inference options. For example, in determining which dose-response curve to choose, the biologically plausible inference options may include the linear, multistage, sublinear, and threshold models. A guideline usually prefers one option, although some guidelines permit the selection of more than one or of all the options. The preferred inference option may be viewed as a default option, i.e., the option chosen on the basis of risk assessment policy that appears to be the best choice in the absence of data to the contrary. A guideline may be said to be flexible according to the degree to which it

allows the default option to be superseded by another inference option as a result of convincing scientific evidence.*

Comparison of IRLG's guidelines with OSHA's guidelines illustrates how comprehensive and detailed guidelines have varied in flexibility. On the issue of benign versus malignant tumors, IRLG's guideline stated:

> The induction of benign neoplasms would, therefore, be considered evidence of carcinogenic activity unless definitive evidence is provided that the test chemical is incapable of inducing malignant neoplasms.

The guideline did not attempt to define the type of definitive evidence that would be needed to demonstrate that a "test chemical is incapable of inducing malignant neoplasms." In contrast, OSHA created strict minimal criteria for acceptance of such evidence:

> (i) Benign tumors. Results based on the induction of benign or malignant tumors, or both, will be used to establish a qualitative inference of carcinogenic hazard to workers. Arguments that substances that induce benign tumors do not present a carcinogenic risk to workers will be considered only if evidence that meets the criteria set forth in 1990.144(e) is provided.

Section 1990.144(e) stated:

> (e) Benign tumors. The Secretary will consider evidence that the substance subject to the rule-making proceeding is capable only of inducing benign tumors in human or experimental animals provided that the evidence for the specific substance meets the following criteria:
> Criteria. (i) Data are available from at least two well-conducted bioassays in each of two species of mammals (or from equivalent evidence in more than two species).

*Flexibility is also intimately related to the legal weight that the agency desires a guideline to have; the implications for flexibility of adopting guidelines under different legal authorities are reviewed in the next section.

> (ii) Each of the bioassays to be considered has been conducted for the full lifetime of the experimental animals.
> (iii) The relevant tissue slides are made available to OSHA or its designee and the diagnoses of the tumors as benign are made by at least one qualified pathologist who has personally examined each of the slides and who provides specific diagnostic criteria and descriptions; and
> (iv) All of the induced tumors must be shown to belong to a type which is known not to progress to malignancy or to be at a benign stage when observed. In the latter case, data must be presented to show that multiple sections of the affected organ(s) were adequately examined to search for invasion of the tumor cells into adjacent tissue, and that multiple sections of other organs were adequately examined to search for tumor metastases.

By leaving open the type of evidence needed to supersede the default option (benign tumors should be considered evidence of carcinogenic activity), IRLG allowed more flexibility than OSHA.

In no case did the IRLG guidelines attempt to restrict the type of evidence that would be needed for acceptance of alternative interpretations. In contrast, OSHA specified minimal criteria for acceptance of alternative interpretations on the issues of negative epidemiologic studies, proof of metabolic differences between animals and humans, and rejection of the use of data from testing at high doses. By invoking such criteria, OSHA attempted to limit the definition of acceptable interpretation and, in so doing, eliminate or reduce scientific debate on controversial issues in its rule-making proceedings.

IRLG also created flexibility by not choosing a default option, i.e., by citing a range of possible inference options to be used in a risk assessment. The statement on interspecies conversion factors illustrates this point:

> Several species-conversion factors should be considered in estimating risk levels for humans from data obtained in another species.

All OSHA guidelines were restricted to the choice of a single inference option.

DEGREE TO WHICH GUIDELINES MAY BE BINDING ON AN AGENCY AND A REGULATED PARTY

The guidelines developed by or for regulatory agencies may vary in their legal status and thus in their procedural implications. For example, OSHA's guidelines (1980) appeared as regulations formally published, after opportunity for public comment, in the Federal Register. In contrast, EPA's guidelines (1976), although eventually printed in the Federal Register, have never been officially subjected to public comment and do not purport to be regulations.

To appreciate the practical differences among the approaches that an agency might follow, it is useful to distinguish three types of administrative documents: regulations (or, synonymously, rules), established procedures (a term we have devised to refer to agency pronouncements that are in some contexts referred to as guidelines), and recommendations. There is no single authoritative definition of the latter two types of document. The discussion here is an attempt to reflect common understanding; it draws as well on the practice, but not the terminology of one agency, FDA (1982a).* An illustration will illuminate the practical differences among these three types of documents. Suppose that an agency decides to adopt, as one of its risk assessment guidelines, the default option that benign tumors should be aggregated with malignant tumors in determining whether a mammalian bioassay demonstrates that an agent causes cancer in the test species. This guideline could be adopted as a regulation, as what we term an established procedure, or simply as a recommendation. For internal purposes, it is not likely to matter which form the

*FDA officially recognizes three types of documents: binding regulations, guidelines, and recommendations. That terminology is potentially confusing here, because we have given guidelines a special meaning, connoting codified principles for risk assessment, that diverges from FDA's legal definition. The reader is referred to the footnote at the beginning of this chapter for a more complete treatment of this discrepancy. We have therefore coined the substitute phrase established procedures, to describe any standards of criteria for fulfilling a regulatory requirement that the agency commits itself to follow until they are formally revoked or revised.

agency's guideline takes. If the guideline is understood to represent prevailing agency policy, the agency's managers can assume that assessors will adhere to it in evaluating test data, regardless of its form. Important differences will be observed, however, in the guideline's impact on interested third parties.

If the guideline were adopted as a regulation, it would be reciprocally binding. Neither the agency nor any private party would be free to argue in a regulatory proceeding that benign and malignant tumors should never be aggregated or should not in a particular instance be aggregated; the agency's regulation would render such arguments legally irrelevant. It is precisely this effect of regulations--i.e., their treatment of previously contested (and in theory still contestable) issues as authoritatively resolved--that OSHA sought when it published its risk assessment guidelines as regulations.

If the guidelines were merely a recommendation, manufacturers of chemicals under evaluation would not be bound by it. They could argue, to the agency or in court, that benign tumors should never be aggregated with malignant tumors or that they should not be aggregated in a particular case. They might not convince the agency, but the agency could not lawfully refuse to consider their arguments or reject evidence supporting them, and they might convince a court that the agency guideline--i.e., its choice of inference options--is wrong generally or inapplicable in a particular case.

If the guideline were an established agency procedure, a private party could similarly argue that it is wrong generally or inapplicable to a particular case. An established procedure does not, therefore, preclude efforts by third parties to treat the benign-versus-malignant issue as an open question. The difference between a recommendation and an established procedure lies in the latter's effect on the agency itself. An agency can depart from a recommendation at any time. Under FDA's practice, however, it may not depart from an established procedure unless it has previously announced that it no longer regards the procedure as sound. In other words, such an established procedure is binding on the agency until formally revoked or changed, and third parties can rely on it and insist that the agency adhere to it.*

*The practical effects of the legal distinctions drawn here are possibly overdrawn. The flexibility accorded by

There is another important difference between regulations and established procedures or, indeed, recommendations. To adopt regulations that have the reciprocally binding effects described above, an agency must follow the procedures prescribed by the Administrative Procedure Act, or by its own statute, for rule-making. At a minimum, these procedures include publication of a proposal, an opportunity for the submission of public comments, and promulgation of a final document that discusses and responds to all significant comments. The process can be long and acrimonious, and that helps to explain why agencies sometimes choose not to adopt policies, particularly those addressing complex issues, in the form of regulations. The same process must be followed to effect changes in regulations once adopted, and that inhibits rapid response to changes in scientific understanding.

ARGUMENTS FOR AND AGAINST THE USE OF GUIDELINES

The advantages and disadvantages listed below constitute an inventory of arguments that have been brought forward by the proponents and critics of guidelines for risk assessment. In most cases, an argument is most convincing for guidelines of a particular form and content, rather than for guidelines in general. For these cases, the characteristics of guidelines that would support or refute an argument are indicated.

---

any set of guidelines depends as much on the language chosen as on the legal form in which they appear. Suppose that an agency's default option is: "Ordinarily benign and malignant tumors shall be equated and their sum used to determine the significance of observed effects, unless (a) new data suggest the inappropriateness of this practice generally, or (b) results from the test in question or other tests of the compound make aggregation inappropriate in the particular case." This text anticipates exceptions, and would not prevent either the agency or a third party from taking a different view about the meaning of a particular test, whether it appeared as a regulation or in some other form.

ADVANTAGES OF GUIDELINE USE

Separation of Risk Assessment from Risk Management

Proponents of guidelines argue that their use would help to separate risk assessment from other parts of the regulatory process. They contend that, when selected inference options are clearly delineated in a formal document, risk assessments will not likely be influenced to fit prior conclusions about regulation of a particular substance. The use of guidelines can also dispel the appearance of such influence when, in fact, there is none. Agencies can defend their assessments on the grounds that they always do them in the way set forth in the guidelines. Compared with reliance on the ad hoc selection of inference options, the use of guidelines could reduce the controversy focused on individual assessments. Debate will shift to the more general discussion of the generic choice of inference options addressed in the guidelines. Guidelines that are comprehensive and detailed will define and bracket the components of risk assessment most completely and explicitly. Thus, such guidelines could probably provide the sharpest distinction between risk assessment and risk management.

Quality Control

Proponents of guidelines argue that their use would ensure the application of selected inference options based on the informed judgment of experts. A single risk assessment requires knowledge in diverse fields, such as epidemiology, biostatistics, toxicology, biochemistry, chemistry, and clinical medicine. Generally, assessors have advanced expertise in no more than a few fields. Guidelines could help to bridge gaps in knowledge by ensuring that decisions are based on judgments formulated by experts in each subject. Guidelines could also help to ensure that assessors apply judgments that are in accord with current scientific thinking in each field. This argument highlights the importance of including experts from a wide range of scientific disciplines in the formulation of guidelines. Furthermore, it suggests that guidelines should be reviewed periodically so that new scientific developments can be accommodated.

Proponents believe that comprehensive, detailed guidelines would be most helpful in providing guidance to

assessors. Comprehensiveness is necessary to provide guidance on all or most of the components of risk assessment. Detailed guidelines could provide an assessor with an expert's insight into aspects of risk assessment that require special consideration. How flexibility could affect quality control is not clear; however, a flexible framework could have a positive effect, especially if guidelines can help an assessor to know when exceptional or novel scientific evidence should be admitted.

Consistency

Almost all guideline documents have stated, in their introductions, that consistency is a major rationale for guideline use. Consistency in risk assessment is important to the agencies, because it helps to ensure fairness and rationality by precluding the arbitrary application of selected inference options that differ from one time to the next. Consistency also permits comparison of risks associated with different chemicals, and this is useful for priority-setting and for facilitating regulatory decision-making. When the same set of guidelines is applied uniformly by all the agencies, government-wide consistency may be improved. This has important implications for interagency coordination and for reducing the possibility that risk assessments by different agencies will be pitted against each other during litigation on a given chemical. Guidelines of a type that fosters consistency among agencies have yet to be adopted and used. In the absence of such guidelines, there are increased opportunities for inconsistency in the choice of inference options available for each risk assessment component and in the conclusions based on those choices. Proponents of guidelines contend it is often difficult even to know whether there is consistency among risk assessments, because of lack of explicit documentation of inference options used.

Comprehensive, detailed guidelines applied uniformly across the agencies appear to be the most suitable form for reducing inconsistency. To ensure thoroughness and clarity in drawing conclusions, assessors should explicitly document the use of such guidelines in their reports. Flexibility does not imply inconsistency in the application of risk assessment policy. The same inference options can be applied consistently, except in instances where convincing contrary scientific evidence is presented. When such evidence is available, the choice of different inference options has a scientific rationale and does not imply an arbitrary shift in risk assessment policy. It is not the same kind of inconsistency as that which can occur when, for example, one assessor uses a species-to-human conversion factor based on surface-area ratios and another, for no better scientific reason, uses a factor based on body-weight ratios.

Predictability

Proponents of guidelines argue that the private sector should be told explicitly which inference options the government will select to evaluate health-effects data. Industry needs this information to assess its own activities and testing programs. Without uniformly applied guidelines, a regulated party may have to call on the agencies for judgments on numerous issues and have no assurance that the judgments will not change unexpectedly or that one agency's judgment will be consistent with another's. Industry representatives have stated their preference for uniform federal guidelines (although they have been much more cautious in discussing the content of and legal weight given to the guidelines). Consider, for example, the following comment by the American Industrial Health Council, regarding the publication of the IRLG cancer guidelines (AIHC, 1979):

> The report is a significant step toward the formulation of a national cancer policy. AIHC supports the report's stated objective of ensuring that regulatory agencies evaluate carcinogenic risks consistently. We strongly urge that this initial step be followed up so that a national cancer policy is developed and conflicting policies among the regulatory agencies are minimized.

This point of view takes on added significance in view of the increasing desire of some states to develop their own cancer policies. Six states have initiated programs thus far, and California has already published its own guidelines (State of California, 1982a,b). For the private sector to have to contend with a range of different policies in different states would clearly be disadvantageous and burdensome. A federal cancer policy could serve as a model to the states and foster a more uniform approach to risk assessment.

Proponents believe that the most useful guidelines in gauging government actions would be detailed and comprehensive. Although flexibility may undermine predictability, it is reasonable to assume that industry would welcome such a trade-off. Guidelines published as established procedures would be the best option, for the regulatory agencies would not change their procedures without formal notice, but the procedures would not be binding on the regulated parties.

Evolutionary Improvement of the Risk Assessment Process

Proponents of guidelines argue that their use provides a locus for debate, examination, and revision of the selected inference options generally used in risk assessment. By contrast, the argument proceeds, when chemicals are evaluated on an ad hoc basis, the focus of debate is shifted from generic issues to case-specific issues, and the methods and assumptions of risk assessment are obscured from critical view.

Over the last decade, new and refined techniques of risk assessment have emerged. Two important examples are the use of short-term in vitro tests to infer carcinogenicity and mutagenicity and the use of dose-response assessment to estimate the magnitude of human risk at low doses. Guidelines may have contributed to the evolution of both by proposing generic interpretations that would be evaluated and tested both in theory and in the laboratory. The choice of a low-dose extrapolation model is a specific example. The first guidelines (FDA, 1973) proposed the use of the Mantel-Bryan model. This choice was the subject of much debate (FDA, 1977, 1979b); newer guidelines have suggested that this model has been discounted by the agencies, in part because it is essentially empirical and lacks biologic relevance with respect to current knowledge about carcinogenesis (IRLG, 1979b; EPA, 1980a). Furthermore, the debate over an appropriate model helped to foster a major research effort. The ED_{01} experiment, also known as the "megamouse study," involved the testing of 24,000 female mice given known carcinogens at low doses in an attempt to determine the shape of the response curve at low doses.

Guidelines that are comprehensive and detailed would invite the most opportunity for debate and evolutionary refinement.

Public Understanding

Because risk assessment is complex, it is easy to parody and demean the process. For example, the decision to label soft drinks containing saccharin was satirized in several highly publicized jokes, e.g., "Caution: Saccharin is hazardous to your rat" and "Drink 800 bottles of pop a day and get cancer." Proponents of guidelines argue that comprehensive, detailed guidelines setting forth the scientific and policy bases of risk assessment could improve public understanding and help to dispel the impression that government actions are based on tenuous and inadequate reasoning.

Administrative Efficiency

Some contend that when risk assessments are performed on a chemical-by-chemical basis without the use of guidelines, too many agency resources are devoted to reargument of the same issues with regulated parties. For example: Should animal carcinogenicity data be used to assess human risk? Should data on animals with a high incidence of spontaneous tumors be considered valid? Should benign tumors be assigned the same weight as malignant ones? Which statistical methods should be applied? Guidelines could reduce repetitious discussion by specifying which types of interpretations are acceptable, given the current state of scientific understanding.

OSHA, in its "Identification, Classification, and Regulation of Potential Carcinogens" (1980), registered concern about its efficiency (only seven rule-making proceedings completed in 9 years) and cited one major reason for its low productivity:

> The necessity to resolve basic scientific policy issues anew, in each rulemaking, has increased the burden on the Department of Labor and members of the scientific community called upon to address these widely accepted policies. Moreover, relitigation of these issues in the federal courts has also drained staff time and energy and has inhibited OSHA initiatives while its policy determinations were repeatedly relitigated.

OSHA maintained that the adoption of cancer guidelines was vital to efficient regulation:

> OSHA believes that this general policy and procedure will facilitate the sifting through the evidence concerning substances which may be interpreted to be potential carcinogens. . . . Without such a system and appropriate criteria, OSHA believes that this task cannot be accomplished in a timely and efficient manner.

Efficiency could best be served by guidelines that are comprehensive, detailed, and inflexible and are adopted as regulations binding on all parties, but this would entail other costs. The disadvantages of such guidelines are described in some of the agruments cited in the following discussion.

DISADVANTAGES OF GUIDELINE USE

Oversimplification

The adoption of guidelines may foster a cookbook approach to risk assessment. The more assessors look at chemicals from a generic point of view, the less they are able to draw distinctions among them on the basis of specific data. The critics' ultimate concern is that blind adherence to guidelines might cause scientific information relevant to a particular chemical to be arbitrarily cast aside because it has not been accommodated in the guidelines.

The following underlined phrases are examples of guidelines that critics believe may lead to oversimplification:

- Use of the most sensitive species to determine risk. Critics contend that, if information shows that metabolic similarity to humans is greater for a species that is less sensitive, data on this species may be preferable.
- Absence of a threshold for carcinogenesis. Critics argue that tumors may be induced by a genetic mechanism or by an epigenetic mechanism. In the latter case, a threshold may exist.
- Unqualified acceptance of positive results at high-dose testing. Critics believe that validity should depend on whether there is a pharmacokinetic difference between high and low dose. Special consideration should be given to whether detoxifying or repair processes are saturated and to whether competing metabolic pathways are involved and become saturated.

Another potential problem is the lack of attention to weighting of evidence. For example, a guideline may simply state that "positive results in animal tests should always outweigh negative results." This does not take into account the quality and statistical power of the different tests; it could foster the attitude that such considerations are of minor importance.

To a large extent, the strength of such criticisms depends on the form and contents of the guidelines. Those which are comprehensive and leave little latitude for exceptional cases tend to maximize the problem of oversimplification; those which are flexible could be most effective in mitigating the problem. In addition, guidelines may explicitly direct the assessor to consider the weight of evidence of a given test result. For example, the IRLG guideline stated that positive results should supersede negative results, but added a caveat: "If the positive result is itself not fully conclusive or if reasons exist for questioning its validity as evidence of carcinogenicity, the result is generally classified as 'inconclusive' or 'only suggestive' even in the absence of other negative results."

Detailed guidelines can reduce the possibility of oversimplification if the intent of detail is to capture for the assessor the complexity of the issue addressed. For example, a guideline might state the scientific basis for the chosen inference option, the kinds of evidence that are typically applicable, circumstances in which acceptance of exceptional evidence may be appropriate, and other rationales for choosing a particular inference option.

Regardless of the form of a guideline, there are some parts of risk assessment, particularly those dealing with the quality of data and the magnitude of uncertainty, that defy or at least resist generic interpretation. Individual judgment is most important in such cases. A guideline should not be viewed as a formula for producing risk assessments without the need for such judgment.

Mixing of Scientific Knowledge and Risk Assessment Policy

Guidelines unavoidably embody both scientific knowledge and risk assessment policy. In the past, regulatory agencies typically used a conservative approach in the development of risk assessment policy, as in the choice of the most sensitive species, use of the most conservative dose-response curve, and the lack of acceptance of negative epidemiologic data. Industry has been highly critical of this approach. Some representatives believe that risk assessment should be solely a scientific function and should be separated from policy decisions. Consider, for example, the American Industrial Health Council's criticism of the IRLG guidelines (AIHC, 1980):

> When the IRLG report speaks of the importance of using conservative methods or assumptions so as not to underestimate human risk, the report is mixing regulatory considerations into the scientific function. The scientific determination should be made separately from the regulatory determinations. On the basis of the best scientific estimate of the real risk, the regulatory agency can then consider costs, benefits and other elements that enter into a regulatory determination.

Furthermore, there is a fear that the mixing will go unrecognized outside the scientific community (AIHC, 1980):

> When value judgments are formalized by the selection, for "conservative" reasons, of a mathematical model or an assumption used for extrapolating human risk, the fact that value judgments have been made escapes the regulator and the public.

The first criticism appears to miss the crucial fact that risk assessment must always include policy, as well as science. The important issues are what the risk assessment policy content is and whether it will be applied consistently or not. The second criticism is most applicable to guidelines that permit an agency to represent as science the conclusions that have been reached in part on the basis of policy considerations. The argument is less applicable to guidelines that explicitly distinguish between scientific knowledge and risk assessment policy

and direct the assessor to address such distinctions when reaching conclusions. Furthermore, it is not clear that risk assessment performed on an ad hoc basis would reduce the opportunity for unrecognized mixing of science and policy; indeed, carefully designed guidelines could help to inhibit such mixing.

Guidelines very different from the kinds described could be designed to be devoid of risk assessment policy choices. They would state the scientifically plausible inference options for each risk assessment component without attempting to select or even suggest a preferred inference option. However, a risk assessment based on such guidelines (containing all the plausible options for perhaps 40 components) could result in such a wide range of risk estimates that the analysis would not be useful to a regulator or to the public. Furthermore, regulators could reach conclusions based on the ad hoc exercise of risk assessment policy decisions.

Misallocation of Agency Resources to Development and Amendment of Guidelines

Critics contend that the dedication of time and resources to the process of guideline development and amendment detracts from an agency's ability to conduct regulatory activities. For example, OSHA's cancer guidelines required 3 years of effort before promulgation of the final rule in January 1980. The full rule-making record eventually exceeded 250,000 pages. OSHA itself offered some 45 witnesses who addressed the scientific content and the policy implications of the proposal, and a much larger number of witnesses appeared in behalf of other participants. The final policy consisted of more than 280 Federal Register pages of preamble and a dozen pages of regulatory text. Notwithstanding this intensive effort, the guidelines have yet to be applied, and new leadership at OSHA is in the process of reevaluating some provisions of the standard.

The procedures required by the Administrative Procedure Act are so elaborate that development and amendment of guidelines written as regulations are expected to demand more intensive effort than guidelines written as established procedures or recommendations. Regardless of the legal status given to the guidelines, their stability over time is susceptible to major changes in policy stances. However, guidelines that clearly distinguish

scientific knowledge from risk assessment policy judgments could provide a locus for facilitating changes in policy orientation. They would define elements of risk assessment policy that are amenable to change and scientific elements that should not be changed for policy reasons. When risk assessment is done on an ad hoc basis, such distinctions may not exist.

Freezing of Science

Critics believe that guidelines would hinder the timely incorporation of important new scientific evidence during standard-setting. The Dow Chemical Company raised this concern about OSHA's cancer guidelines (OSHA, 1980):

> The record . . . has now made it clear that there is absolutely no assurance that the latest scientific evidence in the field will be permitted to be applied under the proposal to any given regulation of a specific chemical substance.

OSHA responded to this criticism by incorporating three amendment procedures into its cancer policy: a general review of the guidelines every 3 years by the directors of the National Cancer Institute, the National Institute of Environmental Health Sciences, and the National Institute for Occupational Safety and Health; recommendations at any time from the National Cancer Institute, the National Institute of Environmental Health Sciences, or the National Institute for Occupational Safety and Health; and petitions from the public. Final amendments would occur only through formal, independent rule-making, to ensure that major changes in the guidelines would not be made during the litigation of individual cases. In industry's perception, the amendment provision did not answer its initial criticism. The American Industrial Health Council characterized the amendment procedures as "a time-consuming and ponderous mechanism for incorporating into the regulatory standards newly available evidence or data concerning heretofore unresolved issues" (OSHA, 1980).

This argument is most applicable to guidelines that are adopted as regulations and to those that are comprehensive and inflexible. When guidelines are flexible and adopted as established procedures or recommendations, the rapid incorporation of novel scientific information is

more easily accommodated. The intent of flexibility is to allow the acceptance of exceptional evidence based on convincing scientific justification. In the case of established procedures or recommendations, changes in guidelines could occur without the necessity of a lengthy rule-making process.

CONCLUSIONS

On the basis of its review of the historical record of guideline development and use and its evaluation of the arguments for and against guideline use, the Committee has drawn several conclusions.

1. All agencies have found it necessary to write guidelines, in part, to make their choice of inference options more evident to the public. However, the application of inference options to specific risk assessments has been marked by a general lack of explicitness.

Because of the lack of explicitness in identifying the choice of inference options in specific risk assessments, it has often been difficult to know whether assessors adhere to guidelines. Within a given program, a consistent set of selected inference options may emerge over time. However, the degree of consistency among programs and agencies is not well defined.

2. Agency guidelines have varied markedly in form and content. Without a deliberate coordinating effort, there is no reason to assume that guidelines will become more nearly uniform.

Although the scientific bases of cancer guidelines developed in the past by the agencies have been generally consistent, the degree to which the guidelines are comprehensive, detailed, flexible, and legally binding has varied widely. EPA's guidelines are statements of broad principles covering a few components in the four steps of risk assessment; they have no regulatory status. OSHA's guidelines were comprehensive and detailed and dealt mainly with hazard identification; they were regulations. CPSC's guidelines were not comprehensive and dealt mainly with hazard identification; they had no regulatory status. FDA's proposed sensitivity-of-method guidelines are comprehensive and detailed for dose-response assessment and exposure assessment; they are regulations. The formation of the IRLG caused the agencies to adopt a single set of

guidelines for the first time, but, since its disbanding in 1981, there has been no further progress on guideline development.*

3. Uniform guidelines for risk assessment (except for exposure assessment) are feasible and desirable.

Guidelines are feasible. The Committee believes that current statutory requirements would not prevent the use of uniform guidelines. Regulators administer laws reflecting social policies that suggest different degrees of acceptable risk. Some argue that uniform guidelines would keep regulators from applying different standards of risk that were based on these laws. However, regulators can apply such standards on the basis of risk management decisions after completion of the risk assessment. Furthermore, feasibility has already been demonstrated by the adoption of the IRLG guidelines.

Uniform guidelines are desirable for several reasons. First, the use of different guidelines by the agencies could undermine the credibility of their risk assessments. Critics of an agency risk assessment might argue persuasively that another agency estimates risk differently, on the basis of a different set of inference options. Second, almost every regulated chemical is in the jurisdiction of two or more agencies, and the possibility of duplication of effort in performing risk assessments on a given chemical could be minimized if the guidelines were applied uniformly. Adoption of uniform guidelines could foster joint risk assessment efforts by agencies interested in regulating the same chemical; or one agency could rely on the assessment of another agency. Through such cooperative efforts, a small agency like CPSC, which lacks the scientific capability of EPA and FDA, could gain help in evaluating complex data. Third, government-wide guidelines could help industry to gauge government actions and to define the types of data and interpretations relevant to industries' own testing programs. Fourth, federal policy could orchestrate efforts toward uniformity among the states.

*The Office of Science and Technology Policy (OSTP), with agency participation, has written a document describing the scientific basis of risk assessment. OSTP envisions the ultimate evolution of a set of principles for risk assessment from this document.

Exposure guidelines, in contrast with guidelines for other risk assessment steps, are not now readily amenable to uniform application in the various agencies. Apart from EPA, the agencies have rather narrowly defined interests regarding exposure, i.e., foods and drugs at FDA, consumer products at CPSC, and occupational hazards at OSHA. Whereas guidelines for the identification of hazard and for the quantitative estimation of risk in test animals may be commonly applied, no such common basis exists for applying exposure assessment guidelines.

4. Even well-designed guidelines may be unsuccessful unless:

- Attention is given to the process by which they are developed.
- They can accommodate change.
- They are viewed as valuable tools, rather than formulas for producing risk assessments.

Because guidelines must include both scientific knowledge and policy judgments, designing a development procedure is a difficult task. Risk assessment requires advanced knowledge in a number of disciplines, and guidelines should be formulated in part on the basis of the best possible scientific expertness in those disciplines. The best mechanism for determining risk assessment policy must be carefully defined. Because of the necessity of considering policy aspects in guidelines, duly appointed public officials must take responsibility for the policy implications. A major goal of the development process should be the assurance that the guidelines preserve a sharp distinction between scientific knowledge and risk assessment policy.

The Committee believes that guidelines should be capable of accommodating evolving scientific concepts in two ways. First, they should be periodically reviewed and, if necessary, revised. Second, they should permit acceptance of new evidence that differs from what was previously perceived as the general case, when scientifically justifiable. However, an unavoidable trade-off results from the use of such flexible guidelines: predictability and consistency may be reduced for the sake of flexibility.

Every risk assessment involves consideration of case-specific factors, such as the quality of the data or the overall strength of the evidence. These factors cannot

be addressed effectively in guidelines. If assessors were to use guidelines in a strictly mechanical fashion, without recognizing the importance of case-specific judgments, the quality of risk assessments could be diminished.

5. Uniform guidelines for effects other than cancer are desirable, but typically they would be based on a less extensive scientific data base.

The same reasons enunciated for the desirability of cancer guidelines impel the conclusion that guidelines are needed to guide assessments of other effects. Scientific data available on these effects may be organized to provide useful information for assessing risk. In fact, guidelines have already been developed for some of these (although never adopted by the agencies), i.e., guidelines for mutagenesis (EPA, 1980; March of Dimes Birth Defects Foundation, 1981) and guidelines for reproductive and teratogenic effects (ORNL/EPA, 1982; March of Dimes Birth Defects Foundation, 1981).

REFERENCES

AIHC (American Industrial Health Council). October 17, 1979. AIHC comments on: A Report of the Interagency Regulatory Liaison Group (IRLG), Work Group on Risk Assessment, entitled "Scientific Basis for Identification of Potential Carcinogens and Estimation of Risks," p. 2.

AIHC (American Industrial Health Council). April 30, 1980. In Review of Public Comments on Statement on Regulation of Chemical Carcinogens; prepared for the U.S. Regulatory Council, Washington, D.C., pp. B-7, B-10.

CPSC (Consumer Product Safety Commission). 1978. Interim policy and procedure for classifying, evaluating, and regulating carcinogens in consumer products. Fed. Reg. 43:25658.

EPA (Environmental Protection Agency). April 5, 1972. Respondents brief in support of proposed findings, conclusions and order at 63-64, in re: Stevens Indus., Inc. (consolidated DDT hearings).

EPA (Environmental Protection Agency). September 5, 1975. Respondents motion to determine whether or not the registrations of mirex should be canceled or amended; Attachment A.

EPA (Environmental Protection Agency). 1976. Health risk and economic impact assessments of suspected carcinogens. Fed. Reg. 41:21402.

EPA (Environmental Protection Agency). 1980a. Mutagenicity risk assessments; proposed guidelines. Fed. Reg. 45:74984.

EPA (Environmental Protection Agency). Office of Water Regulations and Standards. 1980b. Water quality criteria documents; availability. Fed. Reg. 45:79350-79353.

EPA (Environmental Protection Agency). Office of Research and Development, Carcinogen Assessment Group. August 8, 1980c. The Carcinogen Assessment Group's method for determining the unit risk estimate for air pollutants. External review copy, prepared for Office of Air Quality Planning and Standards and Office of Air, Noise and Radiation.

FDA (Food and Drug Administration). 1973. Chemical compounds in food-producing animals; criteria and procedures for evaluating assays for carcinogenic residues. Fed. Reg. 38:19226.

FDA (Food and Drug Administration). 1977. Chemical compounds in food-producing animals; criteria and procedures for evaluating assays for carcinogenic residues. Fed. Reg. 42:10412.

FDA (Food and Drug Administration). 1979a. Assessment of Estimated Risk Resulting from Aflatoxins in Consumer Product Peanut Products and Other Contaminants. Rockville, Md.: Food and Drug Administration.

FDA (Food and Drug Administration). 1979b. Chemical compounds in food-producing animals; criteria and procedures for evaluating assays for carcinogenic residues. Fed. Reg. 44:17070.

FDA (Food and Drug Administration). 1982a. Code of Federal Regulations, Title 21, Section 10.90. Washington, D.C.: U.S. Government Printing Office.

FDA (Food and Drug Administration). 1982b. D & C Green, No. 6, listing as a color additive in externally applied drugs and cosmetics. Fed. Reg. 47:14138.

IRAC (International Agency for Research on Cancer). 1982. General principles for evaluating the carcinogenic risk of chemicals; in IARC monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. IARC, Lyon, France, vol. 29.

IRLG (Interagency Regulatory Liaison Group), Work Group on Risk Assessment. 1979a. Scientific bases for identification of potential carcinogens and estimation of risks. Fed. Reg. 44:39858.

IRLG (Interagency Regulatory Liaison Group), Work Group on Risk Assessment. 1979b. Scientific bases for identification of potential carcinogens and estimation of risks. J. Natl. Cancer Inst. 63:242.

Joint Food and Agricultural Organization and World Health Organization Expert Committee on Pesticide Residues. 1965. Evaluation of the Toxicity of Pesticide Residues in Food: Report of the 2nd Joint Meeting. FAO Meet. Rep. No. PL/1965/10, WHO/Food Add./26.65.

Joint Food and Agricultural Organization and World Health Organization Expert Committee on Food Additives. 1972. Evaluation of Certain Food Additives and the Contaminants Mercury, Lead, and Cadmium. WHO Tech. Rep. Ser. 505. Geneva.

Lehman, A. J., F. A. Vorhes, Jr., L. L. Ramsey, E. C. Hagan, O. G. Fitzhugh, P. J. Schouboe, J. H. Draize, E. I. Goldenthal, W. D'Aguanno, E. T. Lang, E. J. Umberger, J. H. Gass, R. E. Zwickey, K. J. Davis, H. A. Braun, A. A. Nelson, and B. J. Vos. 1959. Appraisal of the Safety of Chemicals in Foods, Drugs and Cosmetics. Association Food and Drug Officials of the United States (AFDOUS).

March of Dimes Birth Defects Foundation. January 26-27, 1981. Guidelines for Studies of Human Populations Exposed to Mutagenic and Reproductive Hazards; Proceedings of Conference, Washington, D.C.

NRC/NAS (National Research Council/National Academy of Sciences), Food Protection Committee, Food and Nutrition Board. 1970. Evaluating the Safety of Food Chemicals. Washington, D.C.: National Academy of Sciences.

ORNL/EPA (Oak Ridge National Laboratory/Environmental Protection Agency). February 1982. Assessment of Risks to Human Reproduction and to Development of the Human Conceptus from Exposure to Environmental Substances. ORNL/EIS-197, EPA/9-82-001.

OSHA (Occupational Safety and Health Administration). 1977. Proposed rule: identification, classification and regulation of potential occupational carcinogens. Fed. Reg. 42:54148.

OSHA (Occupational Safety and Health Administration). 1980. Final rule: identification, classification and regulation of potential occupational carcinogens. Fed. Reg. 45:5001.

OSHA (Occupational Safety and Health Administration). 1981. Identification, classification and regulation of potential occupational carcinogens; conforming deletions. Fed. Reg. 46:4889.

OSHA (Occupational Safety and Health Administration). 1982. Identification, classification and regulation of potential occupational carcinogens. Fed. Reg. 47:187.

OSTP (Office of Science and Technology Policy), Regulatory Work Group on Science and Technology, Executive Office of the President. October 1, 1982. Potential Human Carcinogens: Methods for Identification and Characterization. Part 1: Current Views: Discussion Draft.

Regulatory Council. 1979. Statement on regulation of chemical carcinogens: policy and request for public comment. Fed. Reg. 44:60038.

State of California, Health and Welfare Agency. July 1982a. Carcinogen Identification Policy: A Statement of Science as a Basis of Policy.

State of California, Health and Welfare Agency. October 1982b. Carcinogen Identification Policy: A Statement of Science as a Basis of Policy; Section 2: Methods for Estimating Cancer Risks from Exposures to Carcinogens.

III
Organizational Arrangements for Risk Assessment

The different structures, procedures, and histories of the agencies responsible for regulating toxic substances have produced diversity in their approaches to risk assessment, but common patterns can be discerned, and they permit some broad generalizations about agency organizational arrangements.

First, most agencies have exerted little effort to maintain a sharp organizational separation between employees engaged in assessing the risks associated with substances and those responsible for identifying and evaluating regulatory responses. This is not to suggest that the same persons perform both functions; generally, they do not, for agency organizations reflect considerable specialization, recognizing the distinctive training and capabilities of staff members. However, the two functions are often housed in one organizational unit that is responsible for preparing integrated analyses that incorporate assessments both of risk and of recommended regulatory responses. Sometimes, risk assessment staff are employed in an office that is separate from the office of those who formulate and analyze regulatory options, but, with some notable exceptions, this organizational structure does not lead to a rigid separation of the two staffs.

Second, with the exception of a few experiments in interagency risk assessment during the late 1970s and continuing informal exchanges of information, each agency has performed its own assessments of the risks posed by substances that are candidates for regulation. This operational autonomy does not reflect willful ignorance of the activities of sister agencies or indifference to the desirability of consistency in the evaluation of common candidate substances. Rather, it is a product of

several factors, including the lack of obvious mechanisms for formalized interagency collaboration, the desire of agency policy-makers to reserve authority for policy discretion in reaching conclusions based on risk assessment, the perception that the diversity of types of exposure for which each agency is responsible makes collaborative risk assessment impractical, and differences in regulatory priorities and schedules.

Third, although the four agencies have viewed themselves as ultimately responsible for the risk assessments that support their actions, they often extend their own staff resources available for performing risk assessment by relying on consultants and contractors who are closely supervised by agency personnel. Some agencies--notably the Occupational Safety and Health Administration (OSHA) and the National Institute for Occupational Safety and Health (NIOSH)--whose staffs are small or lack needed expertise rely very heavily on nongovernment contractors and outside scientists in the academic community and government research institutions for performance of risk assessments or specific tasks related to risk assessment (such as literature reviews).

In addition, outside scientists are often called on to review assessments produced by agency staff. Such consultations sometimes take place informally, but often through special advisory committees. These committees can be permanent, such as the Environmental Protection Agency (EPA) Clean Air Science Advisory Committee, or can be created to review particular risk assessments, as is done for many of the Food and Drug Administration (FDA) Bureau of Foods assessments. Some are established by statute, with requirements that they review agency assessments before regulations are proposed. Others are created voluntarily by an agency itself. The members of all federal advisory committees are appointed by the agencies, perhaps with the approval of higher executive-branch authority. Candidates for committee membership usually are identified by agency staff, although some agencies seek nominees from professional organizations and other interested parties. Nominations for some statutorily mandated committees are supplied by an external body, such as the National Academy of Sciences or the National Science Foundation. Advisory panels generally exercise considerable influence and, although legally they are only advisory, share to some extent the agencies' authority to reach conclusions about risk.

TABLE III-1 Examples of Four Models of Organizational Arrangements for Risk Assessment

Integration	Intra-agency Separation	Extra-agency Separation	Scientific Review Panels
OSHA Directorate of Health Standards Programs	EPA Carcinogen Assessment Group[a]	NIOSH-OSHA FDA Drug Evaluation Panels	EPA Science Advisory Board Subcommittee on Airborne Carcinogens
FDA Bureau of Foods		Committees of the National Research Council	
		National Toxicology Program Panel on Formaldehyde[a]	

[a]Separate, centralized assessment body.

TYPES OF ORGANIZATIONAL ARRANGEMENTS

The prominent proposals for reforms in organizational structures and procedures for risk assessment have featured three interrelated principles:

- Risk assessment activities should be strictly separated from the analysis of risk management options and selection of regulatory strategies.
- Risk assessment activities should be centralized in a single body that serves all regulatory agencies.
- Expert panels composed of nonagency scientists should be used either to perform risk assessments for an agency or to review assessments developed by agency staff.

The Committee outlined four idealized models that reflect various combinations of these three principles. The models are integration, intra-agency separation (with or without centralization), extra-agency separation (with or without centralization), and use of scientific review panels. Examples of agency organizations that roughly approximate each model are identified below and in Table III-1. Most of the examples chosen have many distinctive characteristics that obscure or at least outweigh the three organizational principles. In addition, they are not the only examples of a particular model; others could have been reviewed.*

INTEGRATION

In this type of arrangement, a single organizational unit both performs risk assessments and develops regulations. In general, this arrangement is the most common for regulatory programs. For example, for the assessment of chronic hazards involved with chemicals from consumer products, the Consumer Product Safety Commission (CPSC)

*The Committee considered the possible merits of reviewing risk assessment procedures used by other countries as well and decided not to pursue this line of investigation, because of the great differences in political and institutional structures between this country and other countries. Such differences would make it very difficult, if not impossible, to extrapolate findings on institutional structures used in other countries to the United States.

Directorate for Health Sciences is the responsible unit. Before 1977, the Directorate for Health Sciences had few people involved in the risk assessment process, and risk assessments as such were not generally used. Since then, the Directorate has acquired the expertise needed to perform risk assessments itself. The risk assessment is performed within the Directorate, which is distinct from the Commission's politically appointed policy decision-makers. Two different examples of this model examined by the Committee are the OSHA Directorate of Health Standards Programs and the FDA Bureau of Foods (Table III-1). In the former example, risk assessors and those responsible for formulating and recommending regulatory strategies are in the same organizational unit. FDA's Bureau of Foods has a separate office that performs risk assessment, but this separation stems from a functional division of scientific disciplines; it is not intended to and does not result in formal separation of the risk assessment staff from the regulatory staff.

INTRA-AGENCY SEPARATION

In this model, risk assessment is performed by a group that is ostensibly separate from and independent of the office responsible for regulation in the same agency. An intra-agency risk assessment unit could be program-specific or agency-wide. There are examples of program-specific, organizationally separate risk assessment units (notably the Environmental Criteria and Assessment Offices in EPA), but the Committee did not examine them; instead, it reviewed activities of the EPA Carcinogen Assessment Group as an example of an internally separate, agency-wide body.

EXTRA-AGENCY SEPARATION

In this model, an agency's risk assessment is developed outside the agency. The examples reviewed demonstrate the wide variety of arrangements included in this model. Full organizational separation can be achieved by having one institution perform risk assessment and a separate institution regulate exposure to hazardous substances. The relation between NIOSH and OSHA was studied as an example of a permanent, statutory arrangement of this kind. A regulatory agency's use of expert panels to

perform risk assessments can also result in extra-agency separation of risk assessment and regulation. Committees of the National Research Council and several groups of panels used by FDA to review the safety and effectiveness of drugs provide varied examples of such arrangements. The National Toxicology Program Panel on Formaldehyde is an example of an ad hoc assessment group that consisted of government scientists, was organizationally separate from the regulatory agencies (although not without agency members), and served all four agencies (i.e., it was centralized). Because the Interagency Regulatory Liaison Group did not perform risk assessments, it has not been examined as an example of an extra-agency assessment body.

USE OF SCIENTIFIC REVIEW PANELS

Agencies may use independent scientific panels to perform risk assessments or to review assessments prepared by the agencies. This distinction has been used by the Committee to facilitate separate discussion of panels that perform assessments as examples of full organizational separation (see preceding discussion) and panels that review agency assessments as examples of independent review panels. However, the dichotomy is somewhat artificial, in that there may be difficulty in classifying a particular panel. For example, if a panel responsible for performing risk assessments comes to rely heavily on preliminary analyses prepared by agency staff, it can be thought of as acting in a review capacity. Conversely, panels assembled solely for the purpose of reviewing agency assessments have often displayed remarkable independence, sometimes preparing long critiques of agency documents and suggesting substitute findings and reasons. In such cases, to specify which group had performed and which had reviewed the agency's final assessment of risk is difficult.

The extent to which agencies have used independent scientific panels has varied considerably. For example, OSHA has available two types of advisory committees: standing bodies, such as the National Advisory Committee on Occupational Safety and Health, and ad hoc committees that provide advice on specific standards. Members of both types of committee are expected to be knowledgeable about occupational safety and health and may include persons mainly interested in law or regulatory policies. In addition to their professional expertise, however, members of OSHA committees are intended to be represen-

tative of groups interested in occupational health and safety. Several committees have reviewed risk assessments prepared by OSHA or NIOSH. However, because members were intended to be representatives of interest groups, reviews were usually forums for policy debates, not scientific evaluations of risk assessments. In its initial years, OSHA routinely appointed an advisory committee for each regulatory proceeding.

CPSC has had the least experience with expert panels. Before 1981, the Commission was not required to have any assessment of carcinogenic hazard reviewed by an outside panel, although it did make occasional use of such panels (most notably CPSC's request for the National Toxicology Program to form a panel on formaldehyde). CPSC's reauthorization in 1981 included a provision that, before any regulatory action could be proposed on a substance potentially presenting a carcinogenic, teratogenic, or mutagenic hazard, a chronic hazard advisory panel (CHAP) must be established, with the cooperation of the National Academy of Sciences, to review the toxicity of the substance. The first CHAP has recently been convened to review the toxicity of asbestos. Thus, CPSC relies on two methods of peer review for any proposed action. First, independent peer review by outside experts, as well as by a scientific review panel, is performed before a notice of proposed rule-making is issued. Second, the Commission relies on a public rule-making proceeding in accordance with the Administrative Procedure Act during which comment is invited through a Federal Register notice on all aspects of the proposed action. Extensive written comments have been received in the past by this procedure, from industry, consumer groups, members of the academic and scientific communities, and others. Additionally, open, informal public hearings may be held in which interested groups present their views orally; in the past, several such hearings were held during the consideration of a single substance (formaldehyde).

FDA has often used independent scientific panels both to perform and to review agency assessments. The Bureau of Drugs has used standing committees to review and evaluate data on the safety and effectiveness of drug products and to make appropriate recommendations to the Commissioner (see preceding discussion). The use of independent panels by the Bureau of Foods, however, has been on an ad hoc basis, usually at the agency's discretion. However, there are exceptions; for example, the Food, Drug, and Cosmetic Act requires that carcinogenicity

issues related to color additives be referred to a committee of experts selected by the National Academy of Sciences.

EPA, in contrast, has had less choice in its relations with its advisory committees. Several statutes require EPA to consult such committees for scientific review of agency risk assessments or regulations. Examples of mandated advisory committees with a primarily scientific role include the Agency-wide Science Advisory Board; the Clean Air Scientific Advisory Committee, a part of this Board, which reviews criteria documents for air-quality standards; and the Scientific Advisory Panel, which focuses on scientific issues in the Agency's Office of Pesticide Programs. The Committee has examined this panel and a subcommittee of the Science Advisory Board as examples of scientific review panels.

Agency actions, including risk assessments, have been reviewed in the Executive Office of the President; however, because these reviews have, with a few notable exceptions, focused primarily on risk management concerns, the Committee has not examined them.

REVIEW OF AGENCY PROCEDURES FOR RISK ASSESSMENT

This section describes the practices used for risk assessment in each of the organizational examples reviewed by the Committee. The descriptions that follow reveal some strengths and weaknesses of particular approaches and permit some tentative generalizations to be made. Such generalizations, augmented by the experience and judgment of Committee members, lead in turn to recommendations applicable to organizational arrangements for the performance of risk assessment.

The Committee's necessarily retrospective review of agency performance has focused on events and practices of the 1970s, which triggered the current proposals for reform. Changes have been implemented, or at least are contemplated, in the procedures of several of the agencies studied, and the Committee recognized that such changes could alter the performance of risk assessment. Some of the descriptions of agency practices presented here may be dated. However, our purpose is not to describe the current organizational structure of agencies, but rather to discern in the historical record any general relationships between organizational design and procedures and the quality of risk assessments. The

paucity of experience with recent organizational changes and the tendency of any new administration to disclaim the approaches of predecessors while proclaiming the effectiveness of reforms make very recent history less germane to the Committee's purpose.

OSHA'S DIRECTORATE OF HEALTH STANDARDS PROGRAMS (DHSP)

OSHA's health standards were expected by Congress to be based on criteria and recommended standards provided by NIOSH. However, improvements in OSHA's scientific capability and a court directive that OSHA itself review all studies included in the risk assessment supporting a proposed standard prompted the agency to rely less heavily on NIOSH and to begin performing its own risk assessments. Until 1976, OSHA had only a few personnel in the health sciences; however, DHSP has since become an organization staffed primarily by health professionals, including industrial hygienists, responsible for performing risk assessments and for preparing standards, relying on economic and technical analyses supplied by the Office of Regulatory Analysis in a separate directorate (Figure III-1). In addition, the Directorate normally has used a number of consultants who assist with risk assessment or other aspects of standard development, contributing considerable specialized expertise to the organization.

OSHA tried to achieve organizational separation of risk assessment from the preparation of standards in the case of carcinogens. One office in DHSP was supposed to do risk assessment, another to draft standards. In practice, however, such separation was not achieved, largely because personnel shortages required that individual staff members perform both functions.

Agenda and Procedures

DHSP's regulatory and risk assessment agenda has been determined largely by two external forces: petitions by labor unions for action on particular hazards and dramatic discoveries of previously unidentified workplace hazards. Court remands of several OSHA standards, such as the benzene standard, provided new work for OSHA, but none of the mandated re-examinations has led to a final standard. Criteria documents prepared by NIOSH also contributed to OSHA's agenda, in that DHSP staff always read these docu-

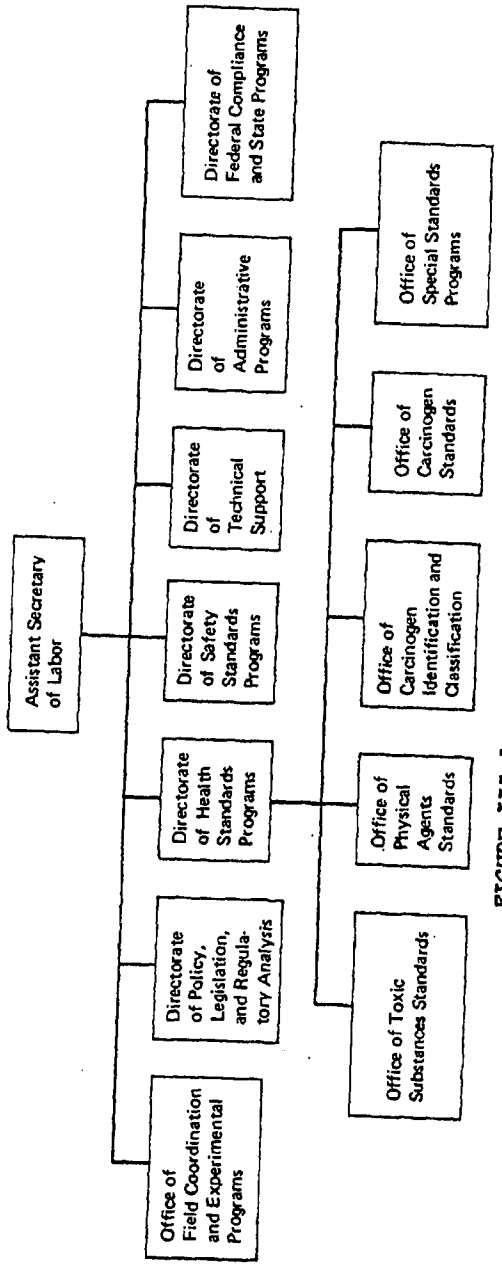

FIGURE III-1 Organization chart of OSHA.

ments when they were received and normally published a Federal Register notice soliciting further information. DHSP's risk assessments usually began with a NIOSH criteria document or other NIOSH input, whatever information was submitted with a labor petition if there was one, the data available from any precipitating discovery, and assessments performed by others, such as the National Academy of Sciences. A literature search and review were conducted by DHSP personnel, often with the help of consultants and NIOSH personnel; and sometimes environmental data on the workplace were solicited or obtained by contractors to contribute to the exposure assessment.

DHSP has not prepared special assessment documents before issuing notices of proposed rule-making. Thus, the first indication provided to the public of the results of an OSHA risk assessment and of the conclusions it intended to draw therefrom was in the Federal Register preamble to its proposed standard. Public comment was invited on all aspects of the proposed standard, including the risk assessment. Extensive written comments were usually received from industry, labor, and others, such as members of the academic scientific community. Customarily a hearing was held at which oral presentations were made and at which questioning of witnesses by OSHA personnel and other witnesses was permitted. The preamble to the final rule, if one were issued, included OSHA's final risk assessment, which incorporated a literature review and OSHA's conclusions on the available scientific data.

In 10 years, OSHA produced permanent health standards for 23 substances or processes, 14 of which were regulated together in a single rule-making. OSHA has also proposed standards for eight substances for which final standards have never been issued, and assessments were conducted for several substances for which new or updated standards are now being considered (Table III-2).

Methods and Use of Guidelines

For most of its history, OSHA has not had formal guidelines for carcinogenic risk assessment. Instead, agency staff have conducted their assessments by choosing options for the components of risk assessment on a case-by-case basis. However, the generic guidelines for identification and classification of carcinogens proposed in 1977 and revised and promulgated in 1980 were intended to

TABLE III-2 A Summary of OSHA Standards

Standards Completed	Standards Proposed, But Not Completed	Standards Being Developed
Asbestos	Arsenic[a]	Ethylene oxide
Vinyl chloride	Beryllium	Asbestos
Arsenic[a]	Sulfur dioxide	Ethylene dibromide
Benzene	Ketones	Cotton dust, nontextile sectors
Coke-oven emission	Hearing conservation (noise)	
14 carcinogens	Toluene	
Lead	Ammonia	
Cotton dust	MOCA	
1,2-Dibromo-3-chloropropane	Trichloroethylene	
Acrylonitrile		

[a]The arsenic standard was remanded to OSHA by the Court of Appeals for the Ninth Circuit for purposes of making a significant-risk determination consistent with the Supreme Court's benzene decision.

replace criteria used in individual cases with generic guidelines that would be applied consistently to all risk assessments of potential carcinogens. The choices incorporated in the 1980 cancer policy reflected the policy orientations of incumbent senior agency officials. Changes now contemplated in these guidelines reflect the policy orientation of the current OSHA management. Similarly, although for many years OSHA did not perform quantitative risk estimates for use in setting standards for carcinogens, it now intends to do so where appropriate. This change results from policy decisions of senior agency officials, based, at least in part, on their interpretation of the Supreme Court's decision on benzene. (Agency officials have interpreted the decision to mean that quantitative dose-response assessments should be

performed for individual substances if data are sufficient.)

Peer Review

OSHA historically has done a less thorough job than other agencies in obtaining relevant scientific information and independent peer review of this information before issuing a notice of proposed rule-making. Instead, the agency has relied primarily on the public rule-making proceeding to identify new information, much of which is in the possession of interested parties and is unlikely to be brought forward except in the context of rule-making. Similarly, although NIOSH's and OSHA's initial assessments often did not provide a critical review of relevant data, critiques of this information were given to the agency during rule-making proceedings, and the agency's final assessment of the risks posed by a chemical often was substantially changed as a result. OSHA's use of rule-making proceedings to provide scientific review stands in sharp contrast with the other agencies' procedures for review. In the Committee's opinion, this reliance on public proceedings to strengthen and refine the scientific basis for the agency's regulatory actions has not been an adequate substitute for independent peer review. In addition, reliance on public proceedings surely precipitated some of the criticism of agency actions and may have jeopardized the scientific integrity and procedural legitimacy of the agency's risk assessments.

Although OSHA's standard-setting actions have stimulated intense controversy, much of it has focused on issues separate from risk assessment. Questions of costs and technologic feasibility (risk management issues) have stimulated much debate. Discussions of the agency's risk assessments have usually focused on its conclusions and their relationship to the agency's regulatory mandate, rather than on its characterization of risk. When OSHA's risk assessments were challenged during rule-making, some key subjects of contention were OSHA's adherence to the assumption that carcinogens have no threshold for causing adverse effects, its tendency to give positive data greater weight than negative data, its use of single epidemiologic studies to support regulatory action, the validity of specific experiments and the agency's interpretation of the data from them, and the decision as to

whether quantitative assessments of risk should be considered. These issues, of course, have both policy and scientific implications.

FDA'S BUREAU OF FOODS

The Food and Drug Administration enforces the Federal Food, Drug, and Cosmetic Act and several related statutes. Its jurisdiction ranges from basic foods to the most advanced pharmaceuticals and medical equipment. The agency assesses the risks associated with thousands of new and existing products every year, functioning through product-oriented units whose responsibilities are reflected in their titles: Foods, Drugs and Biologics, Veterinary Medicine, and Devices and Radiological Health (Figure III-2). The bureaus' agendas are dictated both through internal planning and by external events, particularly applications for approval of new products. Because the Bureau of Foods has had considerable experience with products that pose potential cancer risks, the Committee has focused on this part of FDA in its review.

Agenda and Procedures

The Bureau's risk assessment functions fall into three broad categories: review of petitions for marketing of new compounds for which the manufacturer provides supporting toxicologic and exposure (or use) data; planned retrospective or cyclic review of approved compounds, supporting data on which the Bureau generally must take as it finds them; and review of inadvertent contaminants in food, supporting data on which are derived from many sources, including open scientific literature, monographs, reports, manufacturers' data, and agency-generated data.

In 1981, the Bureau of Foods evaluated 65 food additives, two color additives, and approximately 45 animal-drug petitions. These totals, however, do not reveal the total number of Bureau inquiries that could qualify as risk assessments, albeit perfunctory. Each time a new contaminant is discovered, for example, the Bureau performs some assessment of the risks, although the available data are often limited and little time is available to gather data before it must decide whether to initiate control measures. Similarly, every reported change in

FIGURE III-2 Organization chart of FDA.

degree of contamination invites a new risk assessment. As one would predict, the time and effort required vary with the context. The Bureau's procedures for reviewing food additives, color additives, and residues of animal drugs are more routine than those for evaluating food contaminants, whose occurrence is largely unpredictable. On receipt in the Division of Food and Color Additives, a food-additive petition is evaluated to determine whether it is acceptable for filing. This involves not only review of its formal adequacy, but a preliminary assessment of the toxicologic data to determine whether all potential health effects have been studied.

After official filing of the petition, scientists from the appropriate divisions (ordinarily with the assistance of scientists outside the agency) study the supporting chemical, toxicologic, and exposure data to decide whether the compound is safe. The food-additive law has been construed as requiring, even when the Delaney clause is not applicable, "reasonable certainty" that no consumer will be harmed. No effort is made to evaluate the benefits that an additive might provide, but the Bureau must be satisfied that the additive achieves its intended effects. This exercise usually has two parts: first, Division of Toxicology scientists determine a no-observed-effect concentration for the additive on the basis of acute, subchronic, and chronic feeding studies in animals; second, applying a so-called safety factor, they determine a permissible extent of use in human food or an acceptable daily intake. This value is then compared with the estimated daily human exposure based on the manufacturer's proposed use and predicted human consumption of the foods in which the additive is to be used. An acceptable exposure to an additive is one at which human exposure is at or below the acceptable daily intake. Under current law, this intake value cannot be established for a direct food or color additive that is carcinogenic; such a substance may not be approved for use.

The risk assessment function is performed entirely by Bureau scientists. Bureau staff, including the reviewing scientists, may meet with representatives of the petitioner to discuss uncertainties, request additional data, or suggest reduced use. Typically, both the scientific and the regulatory aspects of food-additive petitions are reviewed and resolved at the division level in the Bureau of Foods. On petitions that raise difficult scientific and policy issues or that pose the question of carcinogenicity, the divisions generally seek advice or direction

from the associate directors, Bureau deputy directors, or the Bureau Director. The Bureau may, in turn, seek advice from the Chief Counsel, from other bureaus, or from the Commissioner's office during the review of petitions that present particular scientific, legal, or policy questions.

Once the responsible unit is satisfied that an additive is approvable and thus that a regulation is appropriate, the Division of Food and Color Additives prepares a document package consisting of an action memorandum, a draft Federal Register document, and supporting material, which is then forwarded through established review channels to the Director's office for final Bureau approval and transmission to the Commissioner's office. The action memorandum recommending approval by the Associate Commissioner for Regulatory Affairs, to whom the Commissioner has delegated formal approval authority, necessarily incorporates both scientific assessments and regulatory judgments. Because the governing legal standard focuses exclusively on the health effects of the additive, the approval process is not influenced by consideration of economic or other benefits.

The sequence of analysis in the Bureau for environmental contaminants does not differ sharply from that described above for food additives, although different divisions may participate in the process and economic factors are consciously considered. The statutory provision under which FDA regulates food contaminants contemplates that it will balance the risk posed by a substance against the effects of reducing consumer exposure, such as loss of food and increases in price. Accordingly, the action memorandum sent to the Bureau Director recommends an exposure limit based on three criteria: an assessment of the risk posed by the contaminant, an evaluation of available methods of chemical analysis to monitor its presence, and an estimate of the economic effects of alternative limits.

Methods and Use of Guidelines

Although the Bureau's approach to the evaluation of acute toxicants has remained stable over a long period, its methods for evaluating potential carcinogens have undergone substantial change since the early 1970s. In 1978, the Bureau Director formed a Cancer Assessment Committee in the Office of Toxicological Sciences to evaluate the carcinogenicity of substances being considered for

approval or regulation and to perform risk assessments. A list of substances reviewed by this Committee in 1981 is given in Table III-3. The 12 members of the Committee are all FDA employees and include toxicologists, pathologists, mathematicians, and chemists. The role of the Committee is to render all final decisions on carcinogenicity for the Bureau of Foods on the basis of scientific information available to it. Its primary function is to determine whether, on the basis of a fair evaluation of all available data, a chemical is a potential or actual carcinogen. Because the Delaney clause, which forbids exposure of any food or color additive that induces cancer, applies to many substances in the Bureau's jurisdiction, quantitative (e.g., dose-response) assessments are not always performed. For some substances, such as contaminants, the magnitude of the risk is relevant, and scientists from the various divisions collaborate with staff responsible for gathering information on human exposure to perform risk characterizations. The Cancer Assessment Committee does not typically prepare formal written assessments, so there is no document available that outlines the relevant data and the rationale for the choices of options made in the assessment of risks. The Cancer Assessment Committee apparently does not follow comprehensive written guidelines, although it does follow some general guidelines that were used in previous decisions and are set out in the agency's drug-residue proposal.

Peer Review

In recent years, the Bureau of Foods has sought independent scientific review of the data on a number of substances. Often Bureau staff informally solicit the judgments of individual outside scientists on major issues. The Bureau routinely uses outside panels established under the auspices of the Federation of American Societies for Experimental Biology for periodic review of substances now generally recognized as safe (GRAS). Ad hoc panels were convened to evaluate the data on such substances as cyclamate, saccharin, Red No. 2, and Red No. 40.

More recently, the Bureau has turned to a standing panel, the Board of Scientific Counselors of the National Toxicology Program. The Board's review of the data on color additive Green No. 5 illustrates the Bureau's approach to external peer review. The Board reviewed the

TABLE III-3 Substances Evaluated for Carcinogenicity by the FDA Cancer Assessment Committee in 1981

Acrylonitrile	1,2-Dichloroethane
Lead acetate	Diethylhexylphthalate
Vinyl chloride	Diethylhexyladipate
Dioxane	Furazolidone
p-Toluidine	Cinnamyl anthranilate
Hydrazine	Trimethylphosphate

original data from a study done by a commercial laboratory, which were submitted with a petition for approval of the substance. The Board also considered aspects of the analysis done by Bureau staff and conducted an independent evaluation of the pathology slides and a statistical analysis of the study results. Bureau scientists asked that the Board reach a conclusion concerning the strength of the evidence of carcinogenicity. Thus, the Board was limited to scientific issues and did not consider the possible social implications of its finding. After the Board's finding that the evidence was inconclusive and before the Bureau's conclusion that the additive was unlikely to be a human carcinogen, Bureau staff performed a risk characterization to estimate the potential risks if this conclusion were in error.

The decision to consult an outside panel for review of risk assessments for potential carcinogens is made by the Chairman of the Cancer Assessment Committee. The Bureau currently is considering establishing a standing committee that could be called on to review agency assessments. It is likely that the impetus to form a standing review committee stems from criticisms of past agency practices, especially those followed in the evaluation of the data on nitrite. In this instance, FDA's contemplated action against nitrite in 1979 was announced before Bureau scientists had had an opportunity to evaluate the critical toxicity data and to refer the data to an independent panel. This controversial chapter in FDA's history of regulating food ingredients has often been cited as demonstrating the need for systematic peer review of the agency's risk analyses in order to avoid the problems that can arise when risk management considerations affect the conduct of risk assessments. The existence of a standing panel, although no guarantee, may discourage

agency officials from deviating from standard Bureau procedures that are now designed to ensure adequate peer review.

EPA'S CARCINOGEN ASSESSMENT GROUP

EPA's Carcinogen Assessment Group (CAG) was created in 1976 by the EPA Administrator to implement generic and uniform agency guidelines for carcinogenic risk assessment. Initially, it was a separate body in the Agency's Office of Research and Development and reported directly to its Assistant Administrator. In 1979, however, the Office of Health and Environmental Assessment was established in the Office of Research and Development, and CAG became one of several assessment groups (Figure III-3). Organizationally, CAG staff are separate from, and independent of, the risk management function; i.e., it is an example of intra-agency separation. It also serves as an example of an internally centralized assessment body, in that it performs assessments for several different regulatory programs in EPA.

Although CAG personnel do meet and talk with regulatory program personnel and are customarily well aware of any programmatic interest in particular substances and of interest-group preferences, this office is insulated from the day-to-day pressures of program offices. Thus, the organizational arrangement that places CAG in the Office of Research and Development does have the initial effect of freeing risk assessment personnel from specific policy issues that arise when risk management options are considered. However, when a scientific review committee examines documents produced by this office later in the process, interest groups are able to express their views and CAG personnel are no longer isolated from such influences.

Currently, all CAG assessments are done by in-house staff, although in the past some were done by consultants. Usually, contractors are employed only for the time-consuming and mechanical task of conducting literature searches. Responsibility for each assessment is assigned to a particular person, but other staff members contribute to various sections according to their particular specialties and expertness. Its staff has been remarkably stable; since 1976, only one person has left the group. As of October 1982, 11 full-time professionals were on its staff, nine of whom had doctorates. Most staff members have an academic background, and their professional work experience averages 10 years. The staff includes